Facultad de Ciencias

Departamento de Química Orgánica

Multicomponent Hydrogen-bonded Macrocyclic Assemblies from a DNA Base Toolkit

Memoria presentada por

CARLOS MONTORO GARCÍA

Para optar al grado de

DOCTOR EN QUÍMICA ORGÁNICA

Madrid, 2017

La presente Tesis Doctoral ha sido realizada en el grupo de *Materiales y Sistemas Moleculares Nanoestructurados* del Departamento de Química Orgánica de la Universidad Autónoma de Madrid bajo la dirección del Prof. David González Rodríguez.

Abbreviations and Acronyms

Common abbreviations and acronyms in organic chemistry have been used following the recommendations published by the American Chemical Society in their "Guidelines for Authors" (*J. Org. Chem.* **2015**).[1] Some other abbreviations have been used and are detailed below:

1D/2D/3D	one-/two-/three-dimensions
A	Acceptor (Hydrogen-bonding/Energy)
A	Adenine/Adenosine[2] – 2-aminoadenine/2-aminoadenosine[3]
AcOEt	ethyl acetate
AFM	Atomic Force Microscopy
BODIPY	boron-dipyrromethene
C	cytosine/cytidine[2]
C	concentration
CB	central block
CCW	counter-clockwise
CD	Circular Dichroism
cor	coronene
CW	clockwise
D	Donor (Hydrogen-bonding/Energy)
DAN	2,7-diamido-1,8-naphthyridine
DAP	diacyldiaminopyridine
DBA	dehydrobenzo[12]annulene
DCvC	Dynamic Covalent Chemistry
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DOSY	Diffusion-Ordered Spectroscopy
EM	Effective Molarity
eq.	equivalent
ESI	electrospray ionization
EXSY	Exchange spectroscopy
FAB	Fast Atom Bombardment
FRET	Förster Resonance Energy Transfer
G	guanine/guanosine[2]
GC-EI	Gas Chromatography-Electron Ionization
h	hours
HRMS	High Resolution-Mass Spectrometry
iC	*iso*cytosine/*iso*cytidine[2]
iG	*iso*guanine/*iso*guanosine[2]

[1] http://pubs.acs.org/paragonplus/submission/joceah/joceah_authguide.pdf
[2] The abbreviations G, C, A, U and T will be used both for nucleobases and their ribonucleosides derivatives.
[3] 2-aminoadenine will be abbreviated as A for the sake of simplicity.

IUPAC	International Union of Pure and Applied Chemistry
K	association constant
M	molar (mol/L)
MALDI	Matrix-Assisted Laser Desorption/Ionization
MS	Mass Spectrometry
NB	nucleobase
NIS	*N*-iodosuccinimide
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
OPE	oligo(phenylene-ethynylene)
P	porphyrin
Pc	phthalocyanine
ppm	parts-per-million
Q-TOF	quadrupole time-of-flight
RNA	ribonucleic acid
rt	room temperature
STM	Scanning Tunneling Microscopy
T	thymine/thymidine[2]
TBAF	tetrabutylammonium fluoride
THF	tetrahydrofuran
TMS	trimethylsilyl
TMSA	trimethylsilylacetylene
T-ROESY	Rotating frame Overhause Effect Spectroscopy
U	uracil/uridine[2]
UG	ureidoguanosine
UHV	ultrahigh vacuum
UPy	ureido pyrimidine
ΔH	change in enthalpy
ΔS	change in enthropy

Table of Contents

Introduction

1. Supramolecular Chemistry and Molecular Self-Assembly

1.1. Non-covalent interactions

Over the last 40 years, since the field of ***supramolecular chemistry*** was founded,[4] chemists have gained considerable knowledge on how to use non-covalent interactions for the synthesis of complex architectures and nanoobjects from "chemically programmed" molecules.[5] This "chemistry beyond the molecule"[6] refers to new supramolecular entities formed by the association of more than one chemical species held together by non-covalent intermolecular forces (Figure 1). This kind of bond is governed by electronic interactions between functional groups in contrast to the covalent bond, where the electrons are shared. Non-covalent interactions are very broad in nature and binding strength (0.4 - 80 kJ/mol).[7] They range from purely electrostatic bonds to dispersion forces, which are governed by enthalpic factors, and to hydrophobic and phase segregation effects, where the entropic contribution may become very relevant. Ion-pair interactions, metal-ligand complexation,[8] halogen[9] and hydrogen (H)-bonding,[10] π-π stacking,[11] or van der Waals interactions[12] are just some of the most popular non-covalent bonds, each of them offering the supramolecular chemist subtle differences in binding strength, kinetics and directionality, as well as in the choice of the most convenient solvent media.

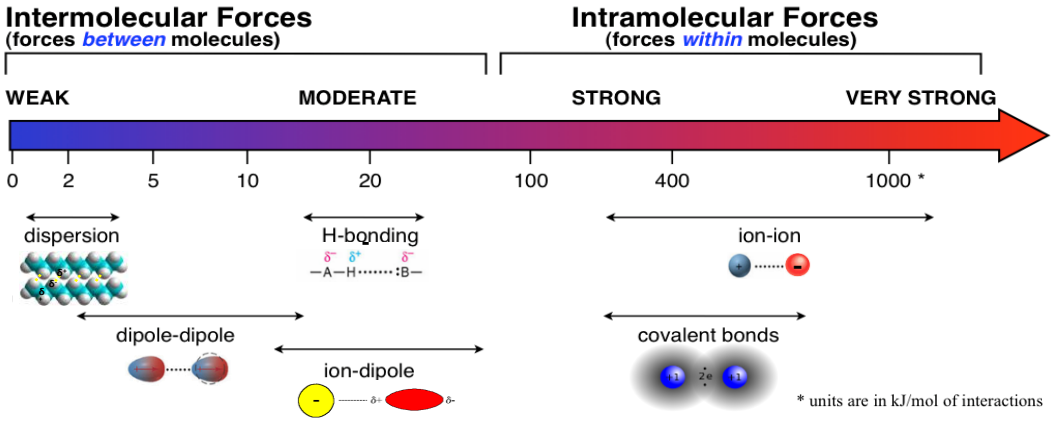


Figure 1. The most important non-covalent interactions along with their typical range of binding strength, in comparison with covalent bonds.

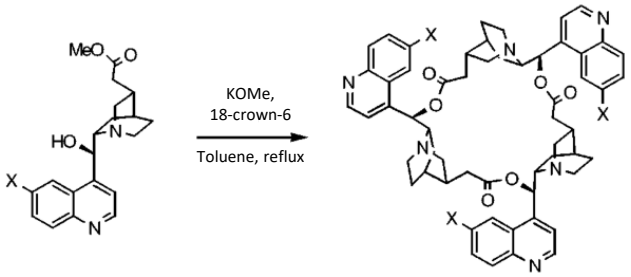
[4] a) J. -M. Lehn, *Angew. Chem. Int. Ed.* **1988**, *27*, 89–112; b) D. J. Cram, *Angew. Chem. Int. Ed.* **1988**, *27*, 1009–1020; c) C. J. Pedersen, *Angew. Chem. Int. Ed.* **1988**, *27*, 1021–1027; d) J. -M. Lehn, *Science*, **2002**, *295*, 2400–2403.
[5] See the special issue on the status of self-assembly at the beginning of the XXI century, *Science*, **2002**, *295*.
[6] J. -M. Lehn, *Chem. Soc. Rev.* **2007**, *36*, 151–160.
[7] H. -J. Schneider, *Angew. Chem. Int. Ed.* **2009**, *48*, 3924–3977.
[8] W. -Y. Sun, M. Yoshizawa, T. Kusukawa, M. Fujita, *Curr. Opin. Chem. Biol.* **2002**, *6*, 757–764.
[9] a) P. Metrangolo, F. Meyer, T. Pilati, G. Resnati, G. Terraneo, *Angew. Chem. Int. Ed.* **2008**, *47*, 6114–6127; b) L. C. Gilday, S. W. Robinson, T. A. Barendt, M. J. Langton, B. R. Mullaney, P. D. Beer, *Chem. Rev*. **2015**, *115*, 7118−7195.
[10] a) P. Schuster, *The Hydrogen Bond-Recent Developments in Theory and Experiments*, Vol. I–III, North-Holland, Amsterdam, **1976**. b) G. A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer, Berlin, **1994**. c) G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford University Press, New York, **1997**. d) G. R. Desiraju, T. Steiner, *The Weak Hydrogen Bond*, Oxford University Press, Oxford, **1999**.
[11] H. Adams, C. A. Hunter, K. R. Lawson, J. Perkins, S. E. Spey, C. J. Urch, J. M. Sanderson, *Chem. Eur. J.* **2001**, *7*, 4863–4877.
[12] K. Mueller-Dethlefs, P. Hobza, *Chem. Rev*. **2000**, *100*, 143–168.

Molecular self-assembly, folding, molecular recognition, host-guest chemistry, mechanically interlocked architectures and dynamic covalent chemistry (DCvC) are just some of the main topics that have been developed in the last decades within the realm of supramolecular chemistry.

1.2. Thermodynamic *versus* Kinetic control

Although there are an increasing number of studies focused on the kinetic control of supramolecular organization, supramolecular chemistry has been classically characterized by a ***thermodynamic control***. Dynamic processes are established that allow the exchange of molecular components until reaching the thermodynamic equilibrium of the system. Therefore, supramolecular chemistry relies on the proper design of basic building blocks to yield the most stable assembled structure. In some cases, *hierarchical self-organization* processes can lead to highly complex matter with a structural control at the nanoscale, which is the basis of the bottom-up approach to nanotechnology (*vide infra*).[13]

DCvC employs reversible covalent bonds instead of noncovalent interactions, and constitutes a magnificent example to illustrate the concept of thermodynamic control. In organic synthesis, kinetically controlled reactions have been used resulting in the irreversible formation of strong covalent bonds (210 - 420 kJ/mol). Once a reaction is finished and the desired product is formed, it is usually not possible transform it in another compound or recover the starting material using the same conditions. In contrast, in DCvC reactions the products obtained depend only on their relative stability, and not on how favourable are the transition states. Hence, DCvC allows some "proof-reading" process where the unstable product is avoided and the desired compound obtained. An interesting example was published by Rowan *et al.*[14] back in 1998, in which they formed cyclic trimers in > 90% yield from cinchona alkaloids, using a transesterification reaction under thermodynamic control. The kinetic cyclization of a similar monomer gave, on the contrary, a complex mixture of cyclic products. Here, the difference in product distribution was attributed to a predisposition of the monomer unit, which adopts a less strained, more stable conformation as a cyclic trimer than as a cyclic tetramer (Scheme1). Likewise, in supramolecular chemistry and molecular self-assembly, thermodynamics govern the structural outcome and allow for some "error checking" of the process, so that the most stable species will be formed.



Scheme 1. Transesterification reaction of cinchona alkaloids in order to form cyclic trimers *via* DCvC.

[13] a) S. J. Rowan, S. J. Cantrill, G. R. L. Cousins, J. K. M. Sanders, J. F. Stoddart, *Angew. Chem. Int. Ed.* **2002**, *41*, 898–952; b) Y. Jin, C. Yu, R. J. Denman, W. Zhang, *Chem. Soc. Rev.*, **2013**, *42*, 6634–6654; c) Y. Jin, Q. Wang, P. Taynton, W. Zhang, *Acc. Chem. Res.* **2014**, *47*, 1575–1586

[14] S. J. Rowan, J. K. M. Sanders, *J. Org. Chem.* **1998**, *63*, 1536–1546.

1.3. Cooperativity and Multivalency

Non-covalent interactions are weak by itself, but they can form much more stable structures through the sum of cooperative and multivalent effects. These effects, which are ubiquitous in biological systems, comprise two distinct but highly interrelated phenomena.

The first one, *cooperativity*, can explain for instance why the binding of one ligand influences a receptor's affinity towards further binding interactions.[15] Cooperativity can be: (1) positive (synergistic) (Figure 2), when the subsequent binding is stronger than the previous one, leading to the growth of the target structure; (2) negative (or interfering), when the first binding event is stronger than the next one; and finally (3) non-cooperative (additive), when all the association constants of the recognition sites are equal.[16]

In self-assembly, cooperativity has a special role that always involves multiple interactions between molecules. Quantifying cooperativity in this way requires the consideration of different factors: (1) effective concentrations[17] of interacting groups within the multivalent receptor or (2) the additivity of free energies.[18] Positive cooperativity intervenes in making the supramolecular assembly progress toward a particular structure by encouraging the formation of a subsequent interaction, after the establishment of the previous one. The Gibbs free energy is more negative than the sum of individual forces, and in this way, the cooperativity of non-covalent interactions plays an important role in the engineering of sophisticated architectures.

However, an interaction between a host/receptor and a guest/ligand that both bear more than one binding site connected through spacers refers to *multivalency*. A single binding site may not enjoy much stability and is easily released by external agents, but many of them together form a strong connection resisting even denaturation environments. By organising the respective binding sites within the molecule, the geometry of supramolecular complexes can be controlled.

Figure 2 depicts different modes of cooperativity that are present both in the natural world and synthetic systems. *Supramolecular aggregation* is a common process used by nature where univalent protein-ligand binding (Figure 2a) is weak in order to achieve a more tight binding (actin fiber).[19] Figure 2b depicts a scheme that can represent the oxygenation of hemoglobin,[20] which is a well-known example of cooperativity in biology. Hemoglobin can bind four oxygen molecules with increasing affinity, until the four binding sites of the protein are occupied, in an intermolecular cooperative process that is known as *allosterism* or *allosteric cooperativity*. Another cooperative process is the *chelate effect* or *chelate cooperativity* (Figure 2c) where an intramolecular binding is preferred over the corresponding intermolecular one and is presented in nature in, for instance, protein folding. The last one is the intermolecular process called *interannular cooperativity*, in which the interaction of the first divalent guest produces a change in the shape and conformation of the host, leading the system to a "frozen" conformation that facilitates the association of a second guest molecule

[15] J. C. Badjic, A. Nelson, S. J. Cantrill, W. B. Turnbull, J. F. Stoddart, *Acc. Chem. Res.* **2005**, *38*, 723–732.
[16] K. A. Connors, *Binding Constants*, Wiley, New York, **1987**.
[17] G. Ercolani, *J. Am. Chem. Soc.* **2003**, *125*, 16097–16103.
[18] P. I. Kitov, D. R. Bundle, *J. Am. Chem. Soc.* **2003**, *125*, 16271–16284.
[19] a) R. T. Lee, Y. C. Lee, *Glycoconjugate J.* **2000**, *17*, 543–551; b) M. Mammen, S. K. Choi, G. M. Whitesides, *Angew. Chem. Int. Ed.* **1998**, *37*, 2755–2794; c) J. J. Lundquist, E. J. Toone, *Chem. Rev.* **2002**, *102*, 555–578.
[20] W. A. Eaton, E. R. Henry, J. Hofrichter, A. Mozzarelli, *Nat. Struct. Bio.* **1999**, *6*, 351–358.

to yield a final rigid assembly. This process is illustrated in nature by self-chaperoning quaternary light-harvesting proteins, as shown in Figure 2d.

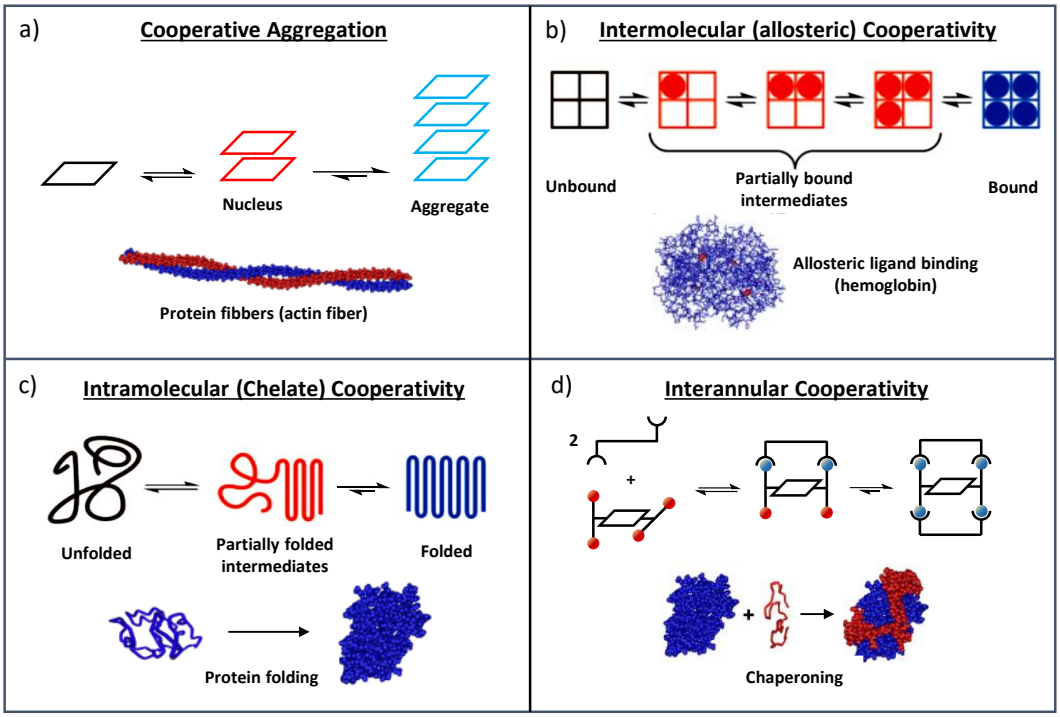


Figure 2. *Representation of supramolecular processes that display positive cooperativity.* (**a**) Cooperative aggregation, (**b**) Intermolecular (allosteric) cooperativity, (**c**) Intramolecular (chelate) cooperativity , and (**d**) Interannular cooperativity.[17,21]

1.4. Self-Sorting Phenomena

Self-sorting[22] in supramolecular chemistry is defined as the high fidelity recognition between molecules (and/or ions) within complex mixtures.[23] If affinity for others is shown, this assembly process is called *self-discrimination* (social self-sorting, which can be, at the same time, integrative or nonintegrative),[24] whereas the affinity for itself would be called *self-recognition* (narcissistic self-sorting)[25] (Figure 3). Self-sorting systems can, in turn, be subdivided into those displaying thermodynamic or kinetic self-sorting, depending if they have reached a thermodynamic equilibrium or can be considered as trapped species under kinetic control, respectively. In general, self-sorting events are directed by the same intermolecular

[21] a) C. A. Hunter, H. L. Anderson, *Angew. Chem. Int. Ed.* **2009**, *48*, 7488–7499.
[22] a) M. M. Safont-Sampere, G. Fernández, F. Würthner, *Chem. Rev.* **2011**, *111*, 5784–5814, b) H. Jędrzejewska, A. Szumna, *Chem. Rev.* **2017**, *117*, 4863–4899.
[23] A. X. Wu, L. Isaacs, *J. Am. Chem. Soc.* **2003**, *125*, 4831–4835.
[24] a) A. Shivanyuk, J. Rebek, Jr., *J. Am. Chem. Soc.* **2002**, *124*, 12074–12075; b) Z. He, W. Jiang, C. A. Schalley, *Chem. Soc. Rev.*, **2015**, *44*, 779–789
[25] P. N. Taylor, H. L. Anderson, *J. Am. Chem. Soc.* **1999**, *121*, 11538–11545.

forces which govern any molecular recognition process, *i.e.* H-bonds,[26] metal-ligand interactions,[27] electrostatic interactions,[28] π-π stacking,[29] and solvophobic effects;[30] and therefore, the factors that determine these recognition events will compromise the *fidelity* of the self-sorting processes. With a good knowledge of noncovalent interactions and playing with the structural characteristics of the monomeric units, we can construct multiple well-defined assemblies from the mixture of different molecules through self-sorting phenomena.

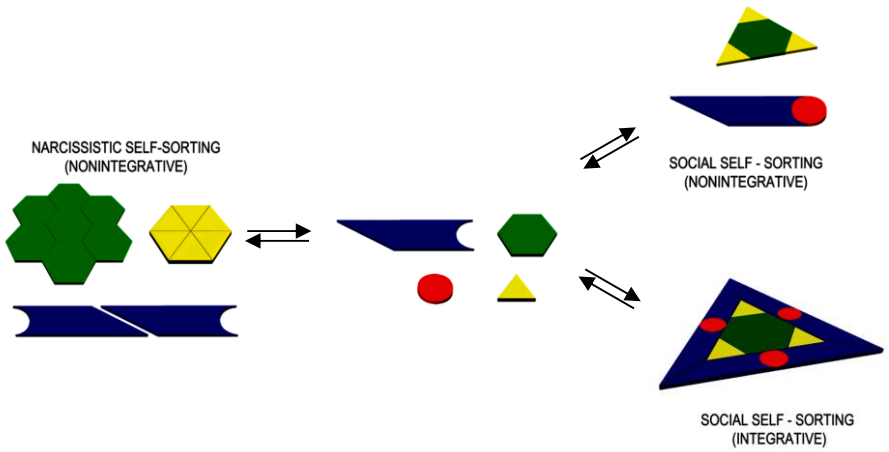


Figure 3. Schematic representation of the different types of self-sorting.[19]

An interesting example is represented by the *meso-meso* linked-pyridine-appended zinc (II) porphyrins possessing axial chirality leading to non-interconvertible *P* and *M* enantiomers, developed by Osuka, Kim and co-workers.[31] In non-coordinating solvents such as CH_2Cl_2, dimeric porphyrins **1a-c** self-assemble into three-dimensional porphyrin boxes **2a-c** (Figure 4), as evidenced by ^{1}H NMR studies and X-ray diffraction of their crystalline structures. These porphyrin boxes form from racemic solutions with extraordinarily large binding constants through simultaneous eight-point metallosupramolecular coordination. Boxes formed from **1a-c** could be demonstrated to be chiral by CD (Circular Dichroism) spectroscopy after resolution of their enantiomers using chiral HPLC. As expected, after chiral resolution, mirror image CD spectra were obtained. The formation of these chiral boxes was attributed to a rather specific self-recognition process of the dimeric porphyrins, which can only form such boxes from identical (that is, single enantiomer) building blocks. Otherwise, upon co-assembly of both enantiomers polymeric structures would form. This last example is most

[26] a) K. A. Jolliffe, P. Timmerman, D. N. Reinhoudt, *Angew. Chem. Int. Ed.* **1999**, *38*, 933–937; b) P. S. Corbin, L. J. Lawless, Z. T. Li, Y. G. Ma, M. J. Witmer, S. C. Zimmerman, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5099–5104; c) Y. G. Ma, S. V. Kolotuchin, S. C. Zimmerman, *J. Am. Chem. Soc.* **2002**, *124*, 13757–13769.
[27] a) D. L. Caulder, K. N. Raymond, *Angew. Chem. Int. Ed.* **1997**, *36*, 1440–1442; b) E. J. Enemark, T. D. P. Stack, *Angew. Chem. Int. Ed.* **1998**, *37*, 932–935; c) R. Stiller, J.-M. Lehn, *Eur. J. Inorg. Chem.* **1998**, 977–982; d) M. Albrecht, M. Schneider, H. Röttele, *Angew. Chem. Int. Ed.* **1999**, *38*, 557–559; e) D. Schultz, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2006**, *45*, 2453–2456; f) Y. M. Legrand, A. van der Lee, M. Barboiu, *Inorg. Chem.* **2007**, *46*, 9540–9547; g) M. Barboiu, F. Dumitru, Y. M. Legrand, E. Petit, A. van der Lee, *Chem. Commun.* **2009**, 2192–2194; h) J. R. Nitschke, *Acc. Chem. Res.* **2006**, *40*, 103–112; i) F. Beuerle, S. Klotzbach, A. Dhara, *Synlett.* **2016**, *27*, 1133–1138.
[28] W. Jiang, H. D. F. Winkler, C. A. Schalley, *J. Am. Chem. Soc.* **2008**, *130*, 13852–13853.
[29] A. D. Shaller, W. Wang, H. Y. Gan, A. D. Q. Li, *Angew. Chem. Int. Ed.* **2008**, *47*, 7705–7709.
[30] a) B. Bilgicer, X. Xing, K. Kumar, *J. Am. Chem. Soc.* **2001**, *123*, 11815–11816; b) N. A. Schnarr, A. J. Kennan, *J. Am. Chem. Soc.* **2003**, *125*, 667–671.
[31] I. W. Hwang, T. Kamada, T. K. Ahn, D. M. Ko, T. Nakamura, A. Tsuda, A. Osuka, *J. Am. Chem. Soc.* **2006**, *128*, 7670–7678.

illustrative for a predefined self-assembly pathway that has been properly encoded in the molecular building blocks to direct a sequence of self-recognition processes. The interplay between enthalpic and entropic contributions obviously directs a high fidelity self-sorting process under the given experimental conditions.

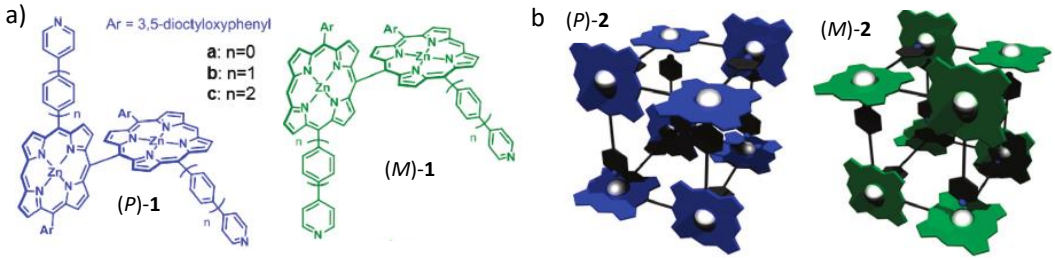


Figure 4. (**a**) Chemical structures of chiral porphyrins **1a-c**. (**b**) Homochiral boxes **2a-c** derived thereof.

1.5. Supramolecular polymerization

Covalent polymers are constituted by the sum of different monomers linked under kinetic conditions due to the fact that the potential barrier for the forward reaction is much lower than the back reaction (depolymerization). In contrast, supramolecular (non-covalent) polymers are defined as: "*polymeric arrays of monomeric units that are brought together by reversible and highly directional secondary interactions, resulting in polymeric properties in dilute and concentrated solution as well as in the bulk. The directionality and strength of the supramolecular bonding are important features of systems that can be regarded as polymers and that behave according to well-established theories of polymer physics".*[32] Such supramolecular polymers can be first classified on the basis of the physical nature of the binding interaction, such as H-bonds, π-π interactions, hydrophobic interactions or metal-ligand binding. Two groups are defined in a second classification scheme, where the first one is represented by a *single* monomer that contains self-complementary or complementary end-group interactions (A-B:A-B:A-B:…), and the second one by *two* bifunctional monomers that contain only one type of interaction (A-A:B-B:A-A:…). Finally a third classification is based on the progress of the Gibbs free energy (ΔG) of the supramolecular polymer as the conversion (p) goes from $p = 0$ (monomer) to $p = 1$ (full conversion). This approach is commonly used to classify the type of supramolecular polymerization mechanisms in three different groups (Figure 5):

(1) ***Isodesmic Polymerization*** (Figure 5a) refers to case where the addition of each monomer to a growing assembly (n-mer) has the same ΔG (or the same association constant; K), meaning that the binding sites of the monomer that form the polymer show the same reactivity regardless of the length of the polymer. When one follows the degree of aggregation as a function of concentration/temperature/solvent nature, sigmoidal curves are obtained that are characteristic of this type of polymerization.

(2) ***Cooperative Polymerization*** (Figure 5b) occurs in two differentiated and consecutive steps. *Nucleation*, the first one, is assumed as a homogenous process that has its own ΔG_n (or K_n). Once a nucleus is formed, the second step is the addition of monomer entities

[32] a) T. F. A. De Greef, M. M. J. Smulders, M. Wolffs, A. P. H. J. Schenning, R. P. Sijbesma, E. W. Meijer, *Chem. Rev.* **2009**, *109*, 5687–5754. b) T. F. A. de Greef, E. W. Meijer, *Nature,* **2008**, *453*, 171–173.

to build the polymer. This process is called *elongation* and is governed by another ΔG_e (or K_e) that, if the process displays positive cooperativity, is lower than the nucleation one and, as in the isodesmic model, is independent of the length of the polymer. Due to the presence of a nucleation step, these kind of polymerizations can enjoy a lower polidispersity than those regulated by isodesmic mechanisms.

(3) ***Ring-Chain Equilibria*** (Figure 5c). If a given ditopic monomer is flexible enough or has a predisposition to cyclize, intramolecular events leading to closed, ring-like systems will compete with the formation of polymer chains. In this scenario, each addition of the respective monomer suppose a new interplay between closed systems or open oligomers. The predisposition of the system to form cyclic species depends on different factors, like steric hindrance within the monomer or the nature and rigidity of the central block linking the noncovalently interacting motifs. The ratio between intra- and intermolecular association constants (K_{intra} and K_{inter}) and the overall concentration will determine if the system will preferably form cyclic or linear systems in solution.

In these supramolecular polymerizations, the average *degree of polymerization* can be controlled by limiting the growth of the chain *via* addition of end-capping units[33] or tuning the nucleation-elongation processes.[31] However, the supramolecular product is most often a mixture of different systems with variable chain lengths.[34]

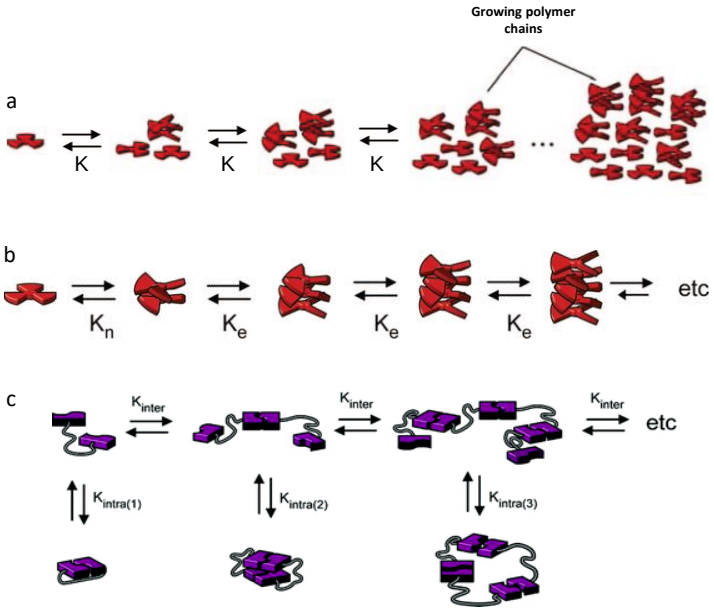


Figure 5. Schematic representation of the supramolecular polymerization of a monomer with two bifunctional sites: (**a**) isodesmic polymerization; (**b**) cooperative polymerization; (**c**) ring-chain mediated polymerization.

[33] a) S. P. Dudek, M. Pouderoijen, R. Abbel, A. P. H. J. Schenning, E. W. Meijer, *J. Am. Chem. Soc.* **2005**, *127*, 11763–11768; b) S. A. Schmid, R. Abbel, A. P. H. J. Schenning, E. W. Meijer, R. P. Sijbesma, L. M. Herz, *J. Am. Chem. Soc.* **2009**, *131*, 17696–17704; c) M. Numata, R. Sakai, *Chem. Lett.* **2014**, *43*, 1890–1892; d) P. Besenius, *J. Polym. Sci., Part A: Polym. Chem.* **2017**, *55*, 34–78.

[34] a) J. S. Moore, *Curr. Opin. Colloid Interface Sci.* **1999**, *4*, 108–116; b) F. W. Zeng, S. C. Zimmerman, S. V. Kolotuchin, D. E. C. Reichert, Y. G. Ma, *Tetrahedron,* **2002**, *58*, 825–843; c) J. -M. Lehn, *Polym. Int.* **2002**, *51*, 825–839; d) U. Michelsen, C. A. Hunter, *Angew. Chem. Int. Ed.* **2000**, *39*, 764–767.

1.6. Non-covalent Synthesis

The term *supramolecular synthesis* or *noncovalent synthesis* is used when a targeted well-defined (uniform) and discrete (monodisperse) structure is reached by mastering the tools of self-assembly.[35] A key term here is *fidelity*, which defines the ability of a given monomer to produce the desired supramolecular assembly over all the possible ones.[36] Due to the complexity of most supramolecular structures, the design of the corresponding monomers must be exquisite. For such a task, the monomers involved in supramolecular growth must present suitable self-assembling directors encoded in their structure, which provide the required chemical information to drive molecular association to a given precise structure, or to structured matter in the nano- and mesoscale. In addition, a knowledge of the individual non-covalent interactions, their synergy and directionality when all of them work together, as well as how the entire system is affected by environmental changes such as solvent polarity, temperature or concentration is demanded.[15,21,37] Furthermore, *orthogonality* between noncovalent interactions, some acting in a given direction and some others in a different one, can be sought in molecular association events for a fine control of the size, shape and function of the desired structure.

The design of "programmed" systems, controlled by molecular information, represents an interesting strategy in materials engineering for the preparation of functional *supramolecular materials*. In general, researchers can follow two different approaches to build a target nanostructure or nanopatterns (Figure 6). A *top-down* approach corresponds to a miniaturization, through the use of existing micro-/nanofabrication and nanopatterning techniques. By means of these techniques one can carve a larger entity into different nanoscale objects/patterns by using externally controlled tools. This method presents some limitations: on one hand, higher accuracies than 100 nm cannot be easily reached and, on the other hand, the level of organization is often not optimal either. In contrast, *bottom-up* approaches are competing with top-down strategies in providing new insights into nanotechnologies. In the bottom-up scenario, one can start from building blocks holding interesting physical and chemical properties and make them grow to form the desired architectures *via* self-organization or self-assembly.[38] Those starting single molecules/ions (monomers) carry the necessary information (size, shape and functionality) to yield a final structure equipped with the desired properties. Consequently, the bottom-up approach deals with increased complexity at the molecular level and maintains a molecule-by-molecule control, where molecular self-assembly and molecular recognition hold an important role.[39]

[35] a) L. J. Prins, D. N. Reinhoudt, P. Timmerman, *Angew. Chem. Int. Ed.* **2001**, *40*, 2382–2426; b) D. N. Reinhoudt, M. Crego-Calama, *Science*, **2002**, *295*, 2403–2407; c) C. Rest, R. Kandanelli, G. Fernández, *Chem. Soc. Rev.* **2015**, *44*, 2543-2572.
[36] E. M. Todd, J. R. Quinn, T. Park, S. C. Zimmerman, *Isr. J. Chem.* **2005**, *45*, 381–389.
[37] a) G. Ercolani, L. Schiaffino, *Angew. Chem. Int. Ed.* **2011**, *50*, 1762–1768; b) Focus Issue on Cooperativity. *Nat. Chem. Biol.* **2008**, *4*, 433–507.
[38] J. V. Barth, G. Costantini, K. Kern, *Nature*, **2005**, *437*, 671–679.
[39] J. K. Gimzewski, C. Joachim, *Science*, **1999**, *283*, 1683–1688.

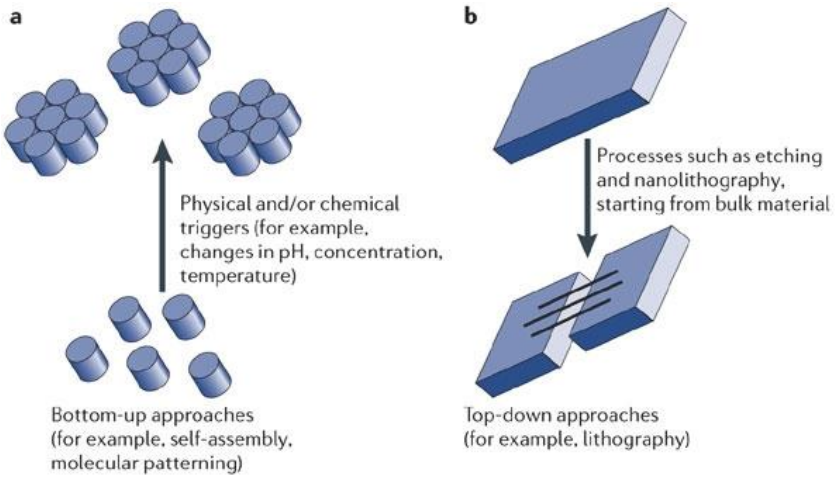


Figure 6. Illustration of the (**a**) "bottom-up" and (**b**) "top-down" approaches.

2. The Hydrogen-Bond

First introduced in 1935 by Bernal and Huggins,[40] H-bonds are electrostatic in nature and occur when a donor (*D*) group with a free acidic hydrogen atom interacts with an acceptor (*A*) group, carrying available non-bonding electron lone pairs. H-bonding is highly directional and selective and provides, therefore, exceptionally controlled geometric and spatial features. H-bonds show a maximum energy value for a 180^{o} angle and, accordingly, a linear binding is preferred. Two of the H-bonding energetic contributions are in the origin of this directionality: the so-called electrostatic or coulomb energy and the charge-transfer energy or covalent bonding. Both have an optimal angle of action. The strength of this interaction depends mainly on the solvent (it can compete for vacant H-bonding sites); the chemical nature of the H-bonding functions (an isolated H-bond is stronger when the basicity of the H-acceptor site and the acidity of the donor site are maximized); and on the total number of H-bonds involved (the forces are additive) and their spatial disposition.[41] Also, the presence of intramolecular H-bonds, tautomerization phenomena, and the electronic effects of substituents can play a decisive role in the final overall strength of this interaction.[42]

2.1. H-bonding Modules

Apart from the number of bonds engaged, the sequence of multiple H-bonds in a particular molecular fragment supposes an important parameter regulating its strength. Indeed, the geometrical disposition of the H-bonding *D* and *A* functionalities is considered as a decisive factor. This phenomenon, with a strong influence in the stability of the H-bond-based complexes, has an electrostatic character. As Jorgensen and co-workers showed,[43] attractive secondary interactions between positively and negatively polarized atoms in adjacent H-bonds increase the H-bonding strength. However, repulsive interactions between two positively or negatively polarized atoms lead to destabilization. As a result of these secondary interactions, the *DD–AA* motif is expected to be more favorable than the *DA–AD* motif.[44] At the same time, for systems containing three hydrogen bonds: the *DDD–AAA* pattern is more stable than the *DDA–AAD*, which in turn is more stable than the *ADA-DAD* pattern. When exploring an example offered by nature, such as DNA base-pairing (Figure 7), one can see that the guanine-cytosine (G–C) pair, with 3 H-bonds, is far stronger than the adenine-thymine one, with only two H-bonds.[45] Even when considering 2-aminoadenine, the binding constant with uracil is significantly lower than between G and C, even if both pairs employ three H-bonds to bind.[40e] In fact, the formation of the G–C dimer (*ADD–DAA* array) involves two attractive a two repulsive secondary interactions, whereas in the 2-aminoadenine–uracil dimer (*DAD–ADA* array), all the secondary interactions are repulsive, resulting in an association constant that is two orders of magnitude lower in $CHCl_3$. Generally speaking, complexes using the *DAD–ADA* motifs exhibit an association constant of around 10^2

[40] a) J. D. Bernal, H. D. Megaw, *Proc. R. Soc. Lond, A.* **1935**, *151*, 384–420; b) M. L. Huggins, *J. Org. Chem.* **1936**, *1*, 407–456.
[41] a) R. P. Sijbesma, E. W. Meijer, *Chem. Commun.* **2003**, 5–16; b) L. Brunsveld, B. J. B. Folmer, E.W. Meijer, R. P. Sijbesma, *Chem. Rev.* **2001**, *101*, 4071–4097; c) J. L. Sessler, C. M. Lawrence, J. Jayawickramarajah, *Chem. Soc. Rev.* **2007**, *36*, 314–325; d) S. Sivakova, S. J. Rowan, *Chem. Soc. Rev.* **2005**, *34*, 9–21; e) S. K. Yang, S. C. Zimmerman, *Isr. J. Chem.* **2013**, *53*, 511–520.
[42] T. Marangoni, D. Bonifazi, *Organic Synthesis and Molecular Engineering*, (Ed.: M. B. Nielsen), John Wiley & Sons, Inc., Hoboken, New Jersey, **2014**, 128–178.
[43] a) W. L. Jorgensen, J. Pranata, *J. Am. Chem. Soc.* **1990**, *112*, 2008–2010; b) J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.
[44] J. Sartorius, H. -J. Schneider, *Chem. Eur. J.* **1996**, *2*, 1446–1452.
[45] A. J. Wilson, *Soft Matter*, **2007**, *3*, 409–425.

M^{-1} in $CHCl_3$, whereas *ADD–DAA* complexes display binding constants of 10^4 M^{-1}.[35,46] *AAA–DDD* arrays can have association constants even higher than 10^5 M^{-1}. Sartorius and Schneider[44] proposed a simple empirical rule to predict the binding strength of a given complex. The free energy for dimerization could be calculated by adding a contribution of 7.87 kJ/mol for each H-bond and ±2.9 kJ/mol for each attractive or repulsive secondary interaction. To sum up, the *DAD–ADA* complex has four repulsive secondary interactions, the *ADD–DAA* complex has two repulsive and two attractive interactions, and the *AAA–DDD* complex has four attracting interactions (Figure 7).

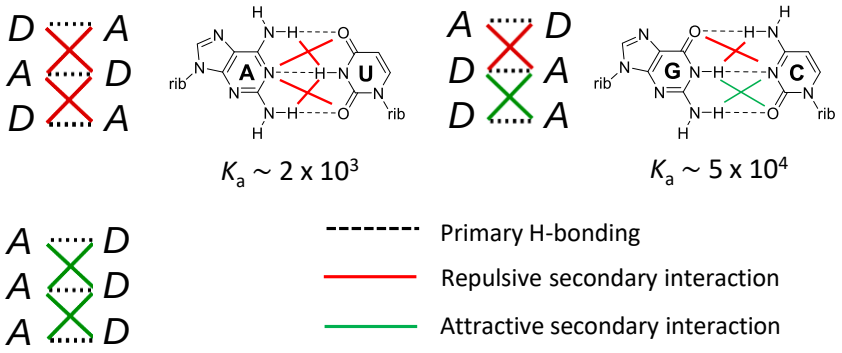


Figure 7. Illustration of secondary H-bonding interactions in triply H-bonded modules.

Many research groups have developed new heterocycles that pair analogously than naturally occurring nucleobases (Figure 8). Additional H-bonds have also been incorporated to different building blocks in order to obtain interesting polymers and supramolecular structures with an increased stability and orthogonality.[40c,47] For instance, the group of Weck utilized cyanuric acid motifs,[48] which formed a three-point H-bonding pattern (*DAD–ADA*); or a six-point H-bonding array or Hamilton wedge (2 x *DAD–ADA*) associated with mono- or ditopic cross-linking agents based on 2,4-diaminotriazine, respectively. Binder and coworkers studied systems with T and diaminotriazine end groups (*ADA–DAD* array), a triple H-bonding motif with four repulsive secondary interactions.[49] Rotello and colleagues prepared functional systems with diacyldiaminopyridine and T[50] or U[51] derivatives, giving rise to a complementary three-point H-bonding (*ADA–DAD*).

[46] T. J. Murray, S. C. Zimmerman, *J. Am. Chem. Soc.* **1992**, *114*, 4010–4011.
[47] a) T. Park, E. M. Todd, S. Nakashima, S. C. Zimmerman, *J. Am. Chem. Soc.* **2005**, *127*, 18133–18142.
[48] K. P. Nair, V. Breedveld, M. Weck, *Macromolecules*, **2008**, *41*, 3429–3438.
[49] F. Herbst, K. Schröter, I. Gunkel, S. Gröger, T. Thurn-Albrecht, J. Balbach, W. H. Binder, *Macromolecules*, **2010**, *43*, 10006–10016.
[50] F. Ilhan, T. H. Galow, M. Gray, G. Clavier, V. M. Rotello, *J. Am. Chem. Soc.* **2000**, *122*, 5895–5896.
[51] U. Drechsler, R. J. Thibault, V. M. Rotello, *Macromolecules*, **2002**, *35*, 9621–9623.

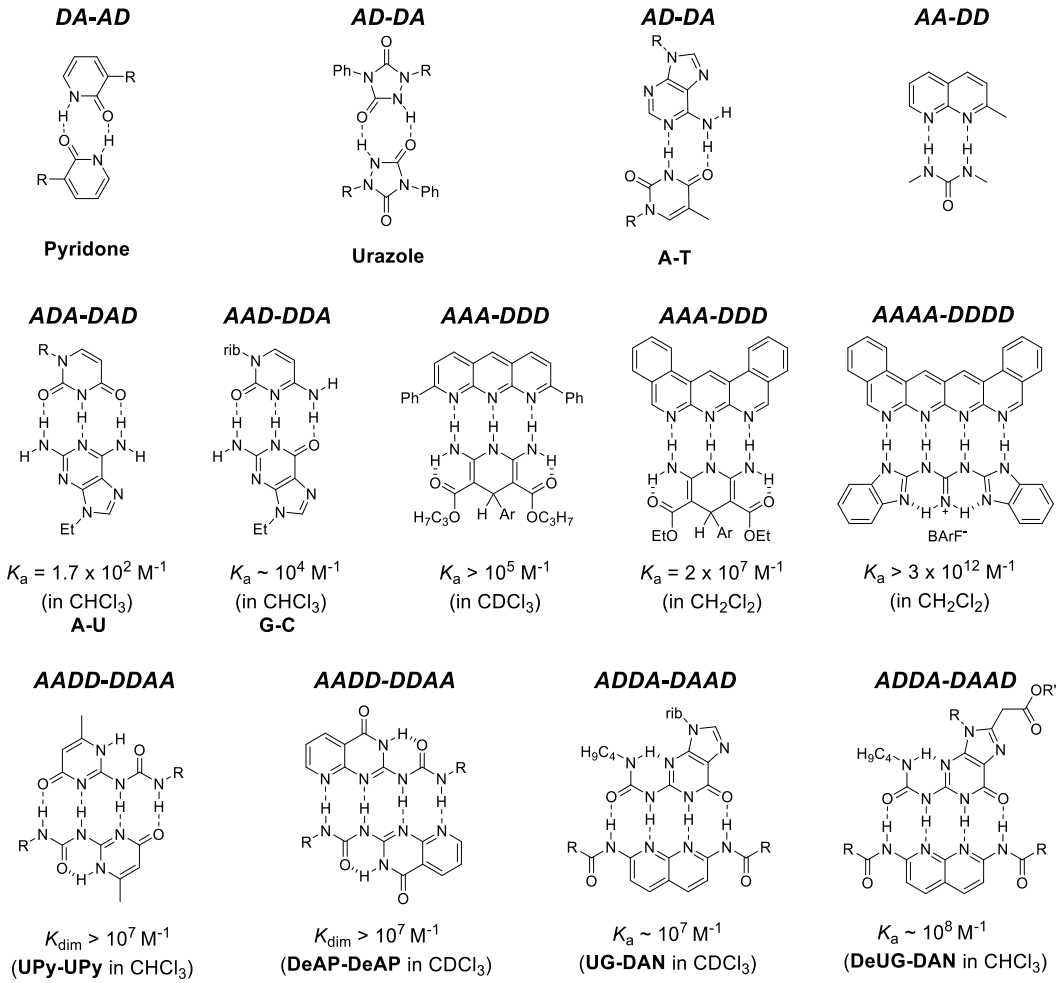


Figure 8. Some representative double, triple, and quadruple H-bonding motifs.

More interestingly, the scientific community has put much effort in developing stronger heterocomplementary arrays of quadruple H-bonds, serving to form supramolecular architectures (Figure 8).[52] The ureido-pyrimidine (UPy) heterocycle is one of the most important and extensively studied H-bonding motifs and a suitable one to engineer nanostructures.[53] UPy shows a strong dimerization in $CHCl_3$[54] as a result of the formation of a self-complementary *DDAA* array of four H-bonds, which are at the same time preorganized by an intramolecular H-bond. Besides UPy, Zimmerman and co-workers used 2,7-diamido-1,8-naphthyridine (DAN), which is a *ADDA* H-bonding array, and ureidoguanosine (UG), which has

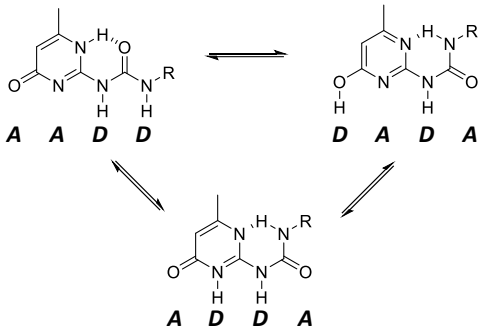
[52] T. Rossow, S. Seiffert, *Supramolecular Polymer Networks and Gels*, (Ed.: S. Seiffert), **2015**, Springer International Publishing AG, Switzerland, 1–46.

[53] a) R. P. Sijbesma, F. H. Beijer, L. Brunsveld, B. J. B. Folmer, J. H. K. K. Hirschberg, R. F. M. Lange, J. K. L. Lowe, E. W. Meijer, *Science*, **1997**, *278*, 1601–1604; b) J. H. K. K. Hirschberg, F. H. Beijer, H. A. van Aert, P. C. M. M. Magusin, R. P. Sijbesma, E. W. Meijer, *Macromolecules*, **1999**, *32*, 2696–2705; c) N. E. Botterhuis, D. J. M. van Beek, G. M. L. van Gemert, A. W. Bosman, R. P. Sijbesma, *J. Polym. Sci., Part A: Polym. Chem.* **2008**, *46*, 3877–3885; d) B. J. B. Folmer, R. P. Sijbesma, R. M. Versteegen, J. A. J. van der Rijt, E. W. Meijer, *Adv. Mater.* **2000**, *12*, 874–878; e) H. Kautz, D. J. M. van Beek, R. P. Sijbesma, E. W. Meijer, *Macromolecules*, **2006**, *39*, 4265–4267; f) L. R. Rieth, R. F. Eaton, G. W. Coates, *Angew. Chem. Int. Ed.* **2001**, *40*, 2153–2156; g) K. E. Feldman, M. J. Kade, E. W. Meijer, C. J. Hawker, E. J. Kramer, *Macromolecules* **2009**, *42*, 9072–9081.

[54] F. H. Beijer, R. P. Sijbesma, H. Kooijman, A. L. Spek, E. W. Meijer, *J. Am. Chem. Soc.* **1998**, *120*, 6761–6769.

a *DAAD* pattern.[55] DAN and UG moieties allow the formation of robust structures, as a result of strong quadruple H-bonding between them, with four repulsive secondary interactions and two attractive ones. UPy can also interact with DAN by quadruple H-bonding in the form of its *ADDA* tautomer (see Scheme 2 below).[53,56] One of the last main publications in this field was described by Leigh and co-workers where an *AAAA-DDDD* quadrupole H-bond array was studied.[57] This new pair revealed exceptionally strong binding for a small-molecule H-bonded complex in a range of different solvents (K_a > 3 x 10^{12} M^{-1} in CH_2Cl_2, 1.5 x 10^6 M^{-1} in CH_3CN and 3.4 x 10^5 M^{-1} in 10% v/v DMSO-d_6/$CHCl_3$) due to the favourable secondary electrostatic interactions between adjacent H-bonds. In a subsequent publication they replaced two of the NH···N H-bonds with CH···N interactions.[58] Surprisingly the association constant in CH_3CN decreased only 2 orders of magnitude, suggesting that the CH···N interactions can be considered part of the *AAAA-DDDD* quadrupole H-bonding array. Furthermore, due to the nature of the complex, the NH···N/CH···N *AAAA-DDDD* motif can be repeatedly switched "on" and "off" in $CDCl_3$ through successive additions of acid and base.

As previously introduced, another parameter that can influence the H-bonding force is the tautomerization process that can occur when heteroaromatic systems are used as molecular recognition units. The possibility to have different tautomeric forms of the same molecule can induce an equilibrium between the *D* and *A* functionalities in the array, modifying their spatial orientation and, therefore, the strength and the specificity of the final interaction. UPy constitutes an example of this phenomenon, as it displays an equilibrium between three forms, where each tautomer exhibits a different H-bonding motif: *AADD*, *DADA* or *DAAD* (Scheme 2). The equilibrium between tautomeric forms, which can be coupled to conformational rearrangements, weakens the association strength and reduces the fidelity of the system.



Scheme 2. UPy tautomeric forms.

[55] a) T. Park, S. C. Zimmerman, S. Nakashima, *J. Am. Chem. Soc.* **2005**, *127*, 6520–6521; b) T. Park, S. C. Zimmerman, *J. Am. Chem. Soc.* **2006**, *128*, 14236–14237; c) T. Park, S. C. Zimmerman, *J. Am. Chem. Soc.* **2006**, *128*, 11582–11590.
[56] X. -Z. Wang, X. -Q. Li, X. -B. Shao, X. Zhao, P. Deng, X. -K. Jiang, Z. -T. Li, Y. -Q. Chen, *Chem. Eur. J.* **2003**, *9*, 2904–2913.
[57] B. A. Blight , C. A. Hunter, D. A. Leigh, H. McNab, Patrick I. T. Thomson, *Nature Chem.* **2011**, *3*, 244–248.
[58] D. A. Leigh, C. C. Robertson, A. M. Z. Slawin, P. I. T. Thomson, *J. Am. Chem. Soc.*,**2013**, *135*, 9939–9943.

2.2. Hydrogen-bonding between nucleobases

Into the natural world, livings systems exhibit a large number of functional designs that suppose a boundless source of inspiration for chemists. By understanding this expertise from biology, scientist are expected to contribute to the solution of many problems in science and technology, in a non-biological context.[59] Much of the inspiration in the supramolecular chemistry field originally came from the double strand assembly found in DNA and RNA. Two antiparallel polynucleotide chains, comprising complementary nucleobase sequences linked to a (deoxy)ribose-phosphate backbone, wind up to generate a supramolecular duplex that is held together by a combination of noncovalent forces.[60] The matching Watson-Crick H-bonding pattern between base pairs is the essential feature that makes two sequence-complementary strands recognize each other and bind more tightly than any other possible supramolecular structure. As mentioned previously, these complementary purine-pyrimidine Watson-Crick pairs are adenine (A)-thymine (T)/uracil (U), which bind through two H-bonds, and guanine (G)-cytosine (C), which associate *via* three H-bonds (Figure 9).

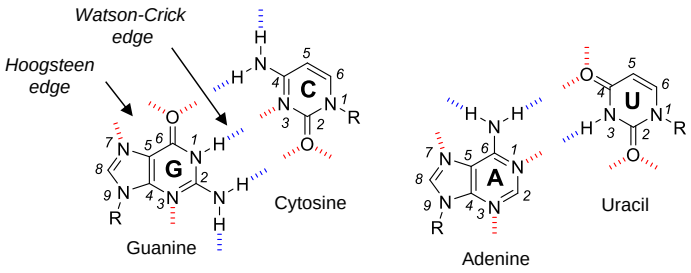


Figure 9. Natural Guanine (G), Cytosine (C), Adenine (A) and Uracil (U) nucleobases. Dashed bonds indicate H-bonding donor (blue) and acceptor (red) sites.

Purines have three different H-bonding interfaces: the Hoogsteen edge, the Watson-Crick edge, and an additional edge comprising N-9, N-3 and (in G) the exocyclic amino group at C-2, which may become competitive specially if N-9 is unsubstituted. Other arrangements of the H-bonding *D* and *A* groups are also possible due to nucleobase tautomerization[61] or ionization, but are far less common. This implies that Watson-Crick pairing is not the only mode of interaction between nucleobases. Actually, there are 28 possible binding motifs that involve at least 2 H-bonds between any pair out of the four common nucleobases (Figure 10).[62] These include reverse Watson-Crick, Hoogsteen and "wobble" (or mismatched) base pairs. In addition to 1:1 pairs, 2:1 complexes (and higher-order ensembles) may also be formed.[63] Natural DNA avoids these non-specific interactions by providing a proper conformational environment for Watson-Crick pairing along the polymeric double strand. However, that does not mean that these secondary interactions are not found in biological media. For instance, DNA can bind diverse molecules and form triplex structures through the Hoogsteen purine

[59] M. Boncheva, G. M. Whitesides, *Biomimetic Approaches to the Design of Functional, Self-Assembling Systems*, Marcel Dekker, Inc., **2004**, 287–294.
[60] E. C. Long, *Fundamentals of nucleic acids*, New York: Oxford University Press. **1996**, 4–10.
[61] Y. P. Wong, *J. Am. Chem. Soc.* **1973**, *95*, 3511–3515.
[62] W. Saenger, *Principles of Nucleic Acid Structures*, Springer, New York, **1984**.
[63] V. Malnuit, M. Duca, R. Benhida, *Org. Biomol. Chem*. **2011**, *9*, 326–336.

edges placed at the grooves.[64] Furthermore, RNA is well-known to use many different types of nucleobase interactions to fold into defined three-dimensional structures.

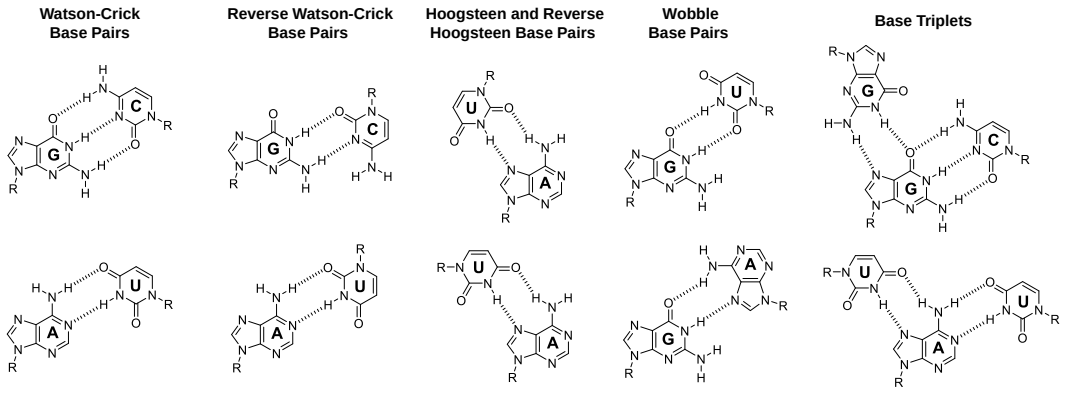


Figure 10. Diverse binding motifs between nucleobases that involve at least 2 H-bonds.

In addition to the formation of hetero-associated pairs, each of the nucleobases can dimerize or oligomerize with diverse association strengths. In particular, G is the base that presents the most versatile supramolecular chemistry, having relatively large dimerization constants (K_a ~ 10^3 M^{-1} in $CHCl_3$)[65] and the highest number of H-bonding sites. Commonly, G derivatives self-associate in apolar solvents into a mixture of oligomeric species by formation of different pairs of self-complementary H-bonds (Figure 11).[66] Linear ribbons may be obtained by interaction between Watson-Crick and Hoogsteen edges (ribbon I) or between the Watson-Crick amide fragment and the aminopyridine fragment (ribbon II).[67] One of these oligomeric species is a cyclic tetramer commonly called *G-quartet*,[68] assembled by DD-AA interactions between Watson-Crick and Hoogsteen edges, whose formation can be templated in the presence of Na^+ or K^+ salts. In these conditions, further stabilization is achieved by G-quartet π-π stacking and cation coordination to the 8 carbonyl groups of sandwiched macrocycles, leading to the so-called *G-quadruplexes*.[65,69]

[64] V. N. Soyfer, V. N. Potaman, *Triple-Helical Nucleic Acids*, Springer, New York, **1995**.
[65] a) J. Sartorius, H.-J. Schneider, *Chem. Eur. J.* **1996**, *2*, 1446–1452.; b) G. Gottarelli, S. Masiero, E. Mezzina, G. P. Spada, P. Mariani, M. Recanatini, *Helv. Chim. Acta*, **1998**, *81*, 2078–2092.
[66] a) J. T. Davis, *Angew. Chem. Int. Ed.* **2004**, *43*, 668–698; b) J. T. Davis, G. P. Spada, *Chem. Soc. Rev.* **2007**, *36*, 296–313.
[67] T. Giorgi, F. Grepioni, I. Manet, P. Mariani, S. Masiero, E. Mezzina, S. Pieraccini, L. Saturni, G. P. Spada, G. Gottarelli, *Chem. Eur. J.* **2002**, *8*, 2143–2152.
[68] a) J. L. Sessler, M. Sathiosatham, K. Doerr, V. Lynch, K. A. Abboud, *Angew. Chem. Int. Ed.* **2000**, *39*, 1300–1303; b) Y. Inui, M. Shiro, S. Fukuzumi, T. Kojima, *Org. Biomol. Chem.* **2013**, *11*, 758–764.
[69] a) D. González-Rodríguez, J. L. J. van Dongen, M. Lutz, A. L. Spek, A. P. H. J. Schenning, E. W. Meijer, *Nature Chem.* **2009**, *1*, 151–155; b) D. González-Rodríguez, P. G. A. Janssen, R. Martín-Rapún, I. De Cat, S. De Feyter, A. P. H. J. Schenning, E. W. Meijer, *J. Am. Chem. Soc.* **2010**, *132*, 4710–4719; c) G. Paragi, C. Fonseca Guerra, *Chem. Eur. J.* **2017**, *23*, 3042–3050.

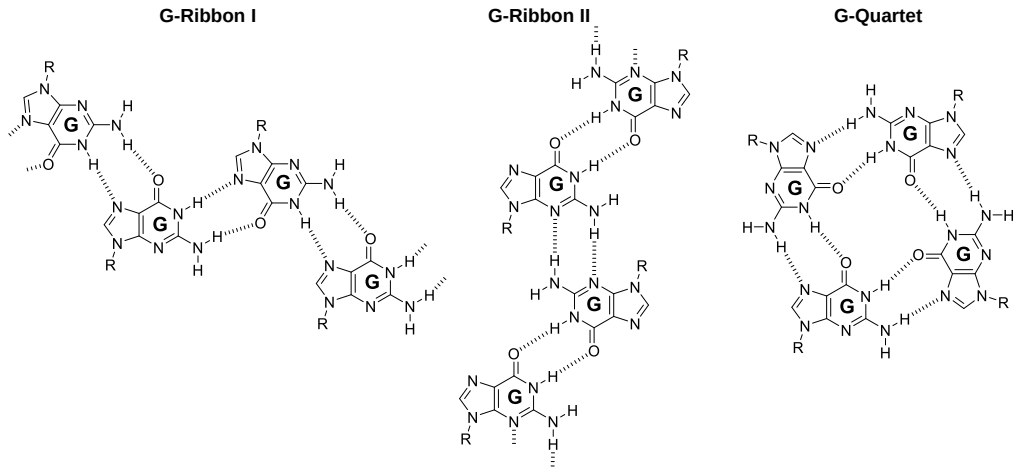


Figure 11. Guanine self-association in different oligomeric species by formation of pairs of self-complementary H-bonds.

In short, the supramolecular versatility of the common purine and pyrimidine nucleobases affords a complex self-assembly playground that chemists are learning to rule. In order to minimize the formation of undesired homo- or hetero-association modes and favor the prevalence of a particular H-bonding mode of choice, the nucleobases allow for a number of synthetic modifications. In some cases, introducing bulky protecting groups and/or blocking potential H-bonding sites on one edge of the nucleobase can be used to guide binding through a different edge. For example, removing N-7 in a purine eliminates Hoogsteen-type associations and enhances Watson-Crick H-bonding. On the contrary, blocking the N-1 in G derivatives favour Hoogsteen interactions. In other cases the H-bonding interfaces can be further extended by incorporating more H-bond donor and/or acceptor sites. For instance, the Hoogsteen face can be expanded by adding other functional groups to the C-8 position of purines.[70]

Finally, other non-natural nucleobase analogues may be produced by exchanging specific functional groups in the purine or pyrimidine heterocycles. The most famous are probably the *iso*guanine (iG) and *iso*cytosine (iC) nucleobases (Figure 12), in which the exocyclic amino and carbonyl groups exchange their positions with respect to G or C.

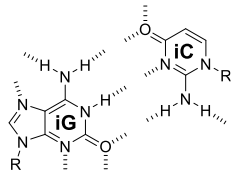


Figure 12. Non-natural H-bonded pairs composed of *isoguanine* (iG) and *isocytosine* (iC).

[70] J. L. Sessler, J. Jayawickramarajah, C. L. Sherman, J. S. Brodbelt, *J. Am. Chem. Soc.* **2004**, *126*, 11460–11461.; b) V. Gubala, J. E. Betancourt, J. M. Rivera, *Org. Lett.* **2004**, *6*, 4735–4738.

3. Noncovalent Macrocyclization Processes

The macrocyclization process is a special case of *ring-chain equilibria*, introduced earlier as a mechanism in supramolecular polymerizations, where a given monomer equipped with at least two binding sites can in principle self-assemble into open (linear) or closed (cyclic) structures. The main difference, and the reason why macrocyclizations are treated here separately, is that we will deal with examples in which a suitable monomer design supplies the required predisposition to self-assemble quantitatively (or close to quantitatively) in a well-defined, discrete macrocyclic structure under a given set of conditions, thus avoiding the formation of polymeric ill-defined species. To achieve this situation, an excellent control on monomer geometry as well as on the type and location of binding sites is essential. The use of templates is another common approach that can lead to ring closure.[71]

Following these requirements, and working under thermodynamic control, the formation of cyclic systems (Figure 13) is highly favoured at the expense of open oligomers since the last intramolecular binding event to form the cycle is highly favoured due to entropic reasons. *Chelate Cooperativity*, introduced earlier as a special case of cooperativity, is the synergistic effect that causes such an increase of stability. This type of cooperativity[21,36] is mainly responsible for many of the "*all-or-nothing*" processes that are characteristic of discrete supramolecular natural and artificial systems. In this case, either a given supramolecular structure is formed in a certain set of conditions or nothing else can survive and the monomeric species is the only entity present.[72] The *Effective Molarity* (*EM*)[73] is the key parameter to quantify the chelate cooperativity, and affords an estimation of how favourable is the intramolecular binding interaction with respect to the intermolecular one. For thermodynamically controlled processes, the *EM* is defined by the ratio between the intramolecular equilibrium constant leading to a cyclic system and the intermolecular one leading to a linear system ($EM = K_{intra}/K_{inter}$).[74] Hence, the product $K_{inter}{\cdot}EM$ provides a measure of the increase in stability when comparing a linear and a cyclic oligomer of a certain length, where K_{inter} considers the additional association to form the cycle and *EM* takes into account that this last binding event is intramolecular.

[71] a) S. R. Seidel, P. J. Stang, *Acc. Chem. Res.* **2002**, *35*, 972–983; b) S. Sato, J. Lida, K. Suzuki, M. Kawano, T. Ozeki, M. Fujita, *Science*, **2006**, *313*, 1273–1276.
[72] J. K. Sprafke, B. Odell, T. D. W. Claridge, H. L. Anderson, *Angew. Chem. Int. Ed.* **2011**, *50*, 5572–5575.
[73] a) A. J. Kirby, *Adv. Phys. Org. Chem.* **1980**, *17*, 183–278; b) L. Mandolini, *Adv. Phys. Org. Chem.* **1986**, *22*, 1–111; c) R. Cacciapaglia, S. Di Stefano, L. Mandolini, *Acc. Chem. Res.* **2004**, *37*, 113-122.
[74] a) X. Chi, A. J. Guerin, R. A. Haycock, C. A. Hunter, L. D. Sarson, *J. Chem. Soc., Chem. Commun.* **1995**, 2563–2565; b) G. Ercolani, *J. Phys. Chem. B*, **1998**, *102*, 5699–5703; c) G. Ercolani, *J. Phys. Chem. B* **2003**, *107*, 5052–5057; d) G. Ercolani, *Struct. Bonding (berlin).* **2006**, *121*, 167–215.

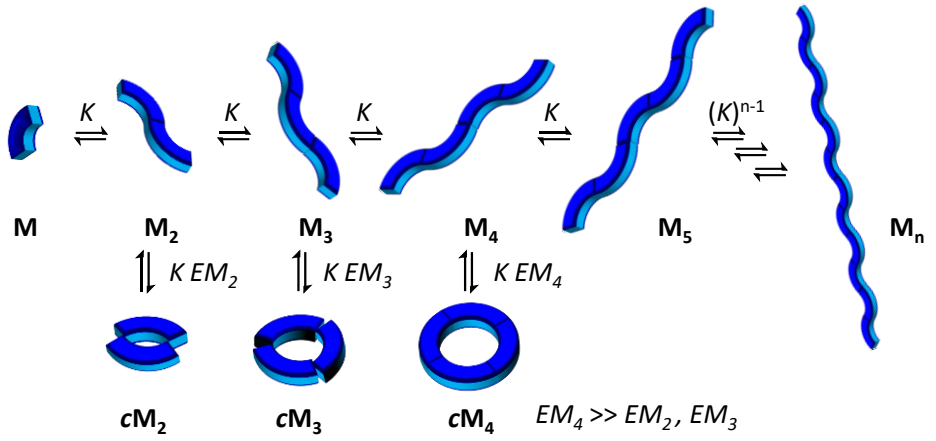


Figure 13. *Supramolecular equilibria of a molecule with a given geometry and two binding sites associating with an intermolecular constant K.* Linear supramolecular oligomers (**M_n**) are in equilibrium with cyclic species (***c*M_n**). In this particular case, a cyclic tetramer is stabilized because the monomer and binding interaction geometric features afford a much higher *EM* value.

Being a thermodynamic magnitude,[75] *EM* has both an enthalpic and an entropic component:

$$EM = e^{-(\Delta H^0_{intra} - \Delta H^0_{inter}/RT)} \cdot e^{(\Delta S^0_{intra} - \Delta S^0_{inter}/R)}$$

The enthalpic component may depend on particular template effects with solvent or specific guest molecules or on electrostatic interactions that affect the cyclic and non-cyclic species in a different way. However, these effects are rare and sometimes difficult to predict, so in most cases this component is solely associated with the strain generated upon ring closure. It is well-recognized that monomers with a preorganized structure that afford unstrained rings are most suited to produce high *EM*s and thus quantitative assembly yields. In the absence of strain, the enthalpic factor becomes negligible and the *EM* only depends on entropic contributions.

The entropic component can reduce the maximum attainable *EM* considerably. It depends on the symmetry and the number of components (*n*) of the cycle, since the reverse ring-opening reaction can take place statistically in *n* sites and because the *EM* tends to dissipate when shared among a relatively large number of molecules. The entropic contribution also decreases with the degrees of conformational freedom that are lost upon cyclization, particularly those related to torsional and rotational motions in the closed *versus* open *n*-mer.[72] It is therefore assumed that the ideal building block must be rigid and the ideal binding interaction must be non-rotatable.

[75] a) C. A. Hunter, M. C. Misuraca, S. M. Turega, *J. Am. Chem. Soc.* **2011**, *133*, 20416–20425; b) M. C. Misuraca, T. Grecu, Z. Freixa, V. Garavini, C. A. Hunter, P. van Leeuwen, M. D. Segarra-Maset, S. M. Turega, *J. Org. Chem.* **2011**, *76*, 2723–2732; c) H. J. Hogben, J. K. Sprafke, M. Hoffmann, M. Pawlicki, H. L. Anderson, *J. Am. Chem. Soc.* **2011**, *133*, 20962–20969; d) C. A. Hunter, M. C. Misuraca, S. M. Turega, *Chem. Sci.* **2012**, *3*, 589–601; e) C. A. Hunter, M. C. Misuraca, S. M. Turega, *Chem. Sci.* **2012**, *3*, 2462–2469; f) H. Adams, E. Chekmeneva, C. A. Hunter, M. C. Misuraca, C. Navarro, S. M. Turega, *J. Am. Chem. Soc.* **2013**, *135*, 1853–1863; g) H. Sun, C. A. Hunter, C. Navarro, S. Turega, *J. Am. Chem. Soc.* **2013**, *135*, 13129–13141.

3.1. Macrocycles Assembled through Metal-Ligand Interactions

The most recurrent interaction to produce cyclic systems and cages is metal-ligand binding between nitrogen heterocycles (mostly pyridines) and transition metal cations (Pd^{2+}, Pt^{2+}, Zn^{2+},..).[8,73d,76] This is a single-point interaction that allows for some degree of conformational flexibility in terms of torsion and rotation around the N-M^{2+} bond.[73d] The extensive use of this interaction, in contrast to the use of H-bonds, and the great development of this field is mainly due to several reasons: (1) metal-ligand bonds are *kinetically more stable* than H-bonds and thus easier to study by NMR techniques; (2) due to this reason and the use of heavy metals, there is a higher probability of obtaining *monocristals* that diffract properly for X-ray analysis; (3) It is a relatively simple, but strong interaction, with *bond energies* that lie between the covalent bond and the classical weak interactions (15 ± 30 kcal.mol^{-1}); (4) it allows for a high *rational design* and *predictability* by simple variation of the size and shape of the building units, or by modifying the geometry of the ligand and the coordination sphere of the metal. Figure 14 provides a beautiful example of the last point: depending on the ligand and metal used, a wide range of discrete supramolecular structures can be obtained.[77]

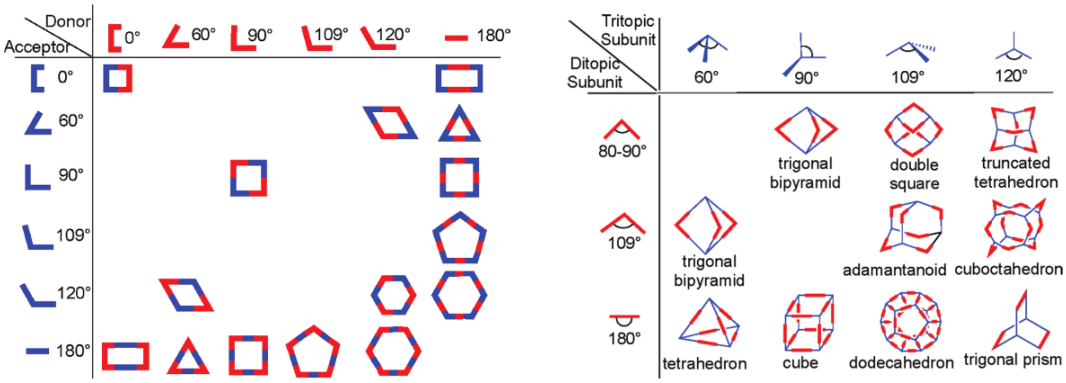


Figure 14. Combination of different building units in order to form a wide range of supramolecular structures.[73a]

Macrocycles

The first square-shaped macrocycle linked through metal-ligand interactions was reported by the group of Fujita in the early 1990 where rigid and linear ligands were combined with metals (Figure 15).[78] The structurally predefined metal corner (Pd or Pt) provides two vacant coordination sites of a 90° angle, while rigid 4,4'-bipyridine ligand possess two lone pairs of 180° angle. These structural features of metal and ligand contribute convergently to the formation of a square framework at room temperature in water. In contrast to Pd (II) that produced the square shape macrocycle quantitatively, Pt (II) gave initially the kinetically distributed oligomeric products, which could be converted in the square species by heating at 100° for more than four weeks. An even more effective route to palladium and platinum-

[76] a) M. Fujita, M. Tominaga, A. Hori, B. Therrien, *Acc. Chem. Res.* **2005**, *38*, 371–380; b) S. Leininger, B. Olenyuk, P. J. Stang, *Chem. Rev.* **2000**, *100,* 853–908; c) F. Würthner, C.-C. You, C. R. Saha-Möller, *Chem. Soc. Rev.* **2004**, *33*, 133–146.
[77] a) R. Chakrabarty, P. S. Mukherjee, P. J. Stang, *Chem. Soc. Rev.* **2011**, *111*, 6810–6918; b) M. M. J Smulders, I. A. Riddell, C. Browe, J. R. Nitschke, *Chem. Soc. Rev.* **2013**, *42*, 1728–1754; c) T. R. Cook, J. Stang, *Chem. Rev.*, **2015**, *115*, 7001–7045.
[78] a) M. Fujita, J. Yazaki, K. Ogura, *J. Am. Chem. Soc.* **1990**, *112*, 5645–5647; b) M. Fujita, J. Yazaki, K. Ogura, *Chem. Lett.* **1991**, 1031–1032.

containing molecular squares was developed by Stang *et al.* by employing metal corners with chelating bisphosphino ligands that are well-soluble in organic media.[79]

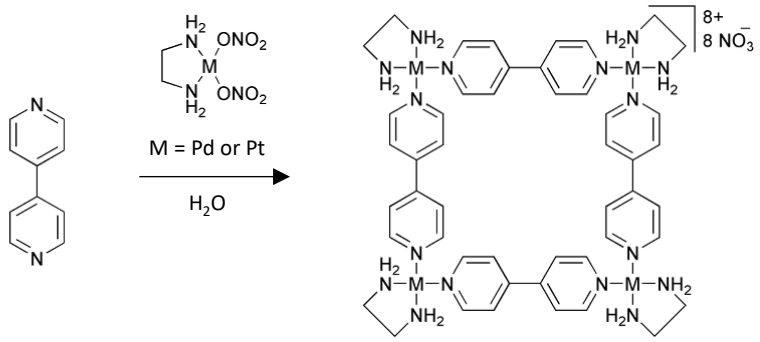


Figure 15. Example of the first metallosupramolecular square constructed by Fujita and co-workers.

Another approach is the formation of tetra-porphyrin squares from properly substituted zinc porphyrins through self-complementary coordinative interactions.[80] Hunter *et al.* have reported on the construction of a molecular square from a zinc porphyrin bearing a *p*-(iso-nicotinamide)phenyl group. In the monomeric structure, the angle between the Lewis-acidic coordination site of the zinc porphyrin and the Lewis-basic lone pair of the pyridine ligand is approximately 90°. Thus, a tetrameric macrocycle was formed through the intermolecular axial coordination of the pyridyl moiety to the zinc porphyrin in a head-to-tail fashion (Figure 16a). Osuka and colleagues constructed a more rigid tetrameric square that could be isolated in the solid state and its structure was unequivocally confirmed by X-ray analysis (Figure 16b).[79b] The *EM* values of the complexes showed in Figure 16 were calculated by Ercolani,[74d] which are 0.6-0.9 molL^{-1} and 18-27 molL^{-1} respectively.

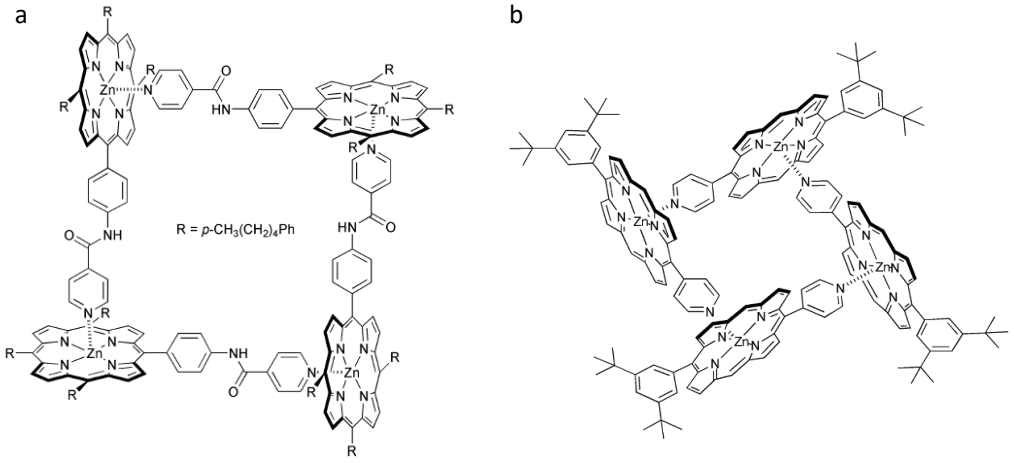


Figure 16. Example of the first metallosupramolecular square constructed by Fujita and co-workers.

[79] P. J. Stang, B. Olenyuk, *Acc. Chem. Res*. **1997**, *30*, 502–518 and references therein.
[80] a) X. Chi, A. J. Guerin, R. A. Haycock, C. A. Hunter, L. D. Sarson, *J. Chem. Soc., Chem. Commun.* **1995**, 2567–2569; b) A. Tsuda, T. Nakamura, S. Sakamoto, K. Yamaguchi, A. Osuka, *Angew. Chem. Int. Ed.* **2002**, *41*, 2817–2821.

3D structures: Cages and Prisms

The ability of Pd(II) to support a homoleptic environment of four pyridyl donors is the basis for many of *hollow spherical constructions*, like the ones produced by Fujita's group. The simplest of these M_nL_{2n} designs uses two metal centers bridged by four ligands. Using this approach, Fujita and co-workers have developed a rich library of spheres exhibiting interesting chemistry, including cubes (M_6L_{12}), cuboctahedra ($M_{12}L_{24}$), and rhombicuboctahedra ($M_{24}L_{48}$) and related species. The synthesis, characterization and chemistry of these self-assembled macrocycles, has been summarized recently.[81]

The first cuboctahedron, which is a self-assembled macrocycle from 36 components and the most widely studied, was first reported in 2004.[82] This highly spherical structure is constituted by simple ditopic bis(pyridine) ligands upon reaction with Pd(II) ions (Figure 17a). The simplicity and the downfield shifts of the pyridine proton signal in the ^{1}H NMR spectra suggests the formation of a well-defined and symmetrical sphere in solution due to the ideal angle of 120° adopted by corresponding furan ligands. The diameter of the sphere was estimated by DOSY and STM studies to be around 3.5 nm, and the final confirmation of the $M_{12}L_{24}$ distribution was afforded by cold-spray ionisation mass spectrometry (CSI-MS) and X-ray structural determination. The stability of this kind of complexes comes from the cooperativity of 48 weak π-Pd(II) interactions, and ligand exchange occurs only very slowly (*ca.* 20 days).

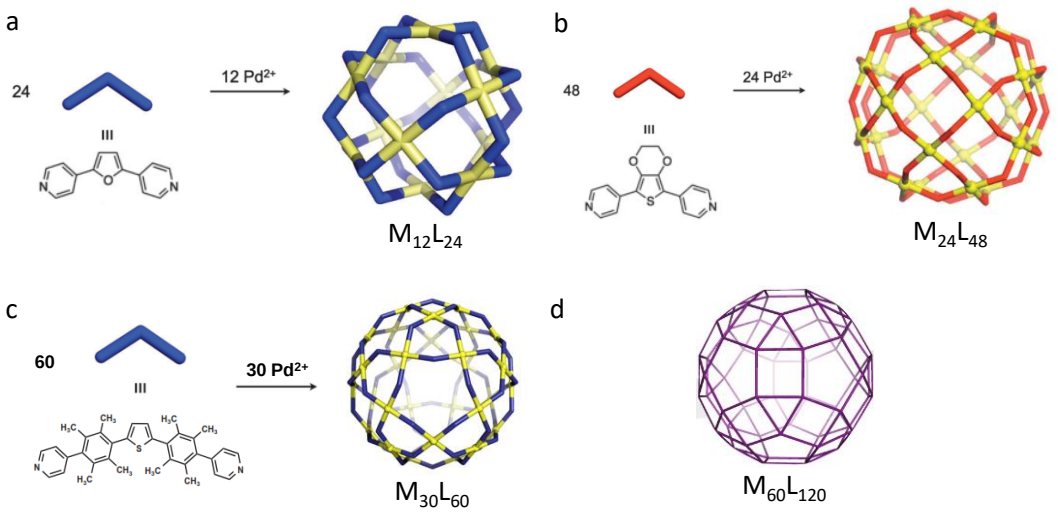


Figure 17. Different selected cages developed by Fujita and co-workers trough metal-ligand interactions.

Now, in order to expand the formation of supramolecular boxes in solution, it is necessary to assume the perfectly square planar metal centres and ligands with a bend angle of 135° to form rhombicuboctahdedrons with high fidelity.[83] A thiophene-based ligand is used in this case for the formation of a 72-component ($M_{24}L_{48}$) rhombicuboctahedron (Figure 17b),

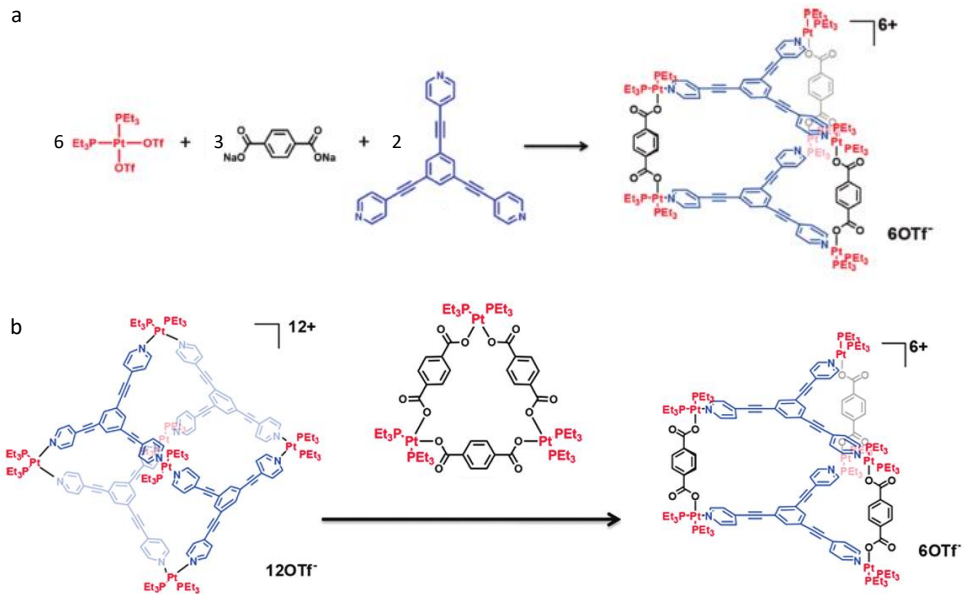
[81] K. Harris, D. Fujita, M. Fujita, *Chem. Commun*. **2013**, *49*, 6703-6712.
[82] M. Tominaga, K. Suzuki, M. Kawano, T. Kusukawa, T. Ozeki, S. Sakamoto, K. Yamaguchi and M. Fujita, *Angew. Chem. Int. Ed.*, **2004**, *43*, 5621–5625.
[83] Q. F. Sun, J. Iwasa, D. Ogawa, Y. Ishido, S. Sato, T. Ozeki, Y. Sei, K. Yamaguchi and M. Fujita, *Science*, **2010**, 328, 1144–1147.

which was corroborated by X-ray, CSI-MS and DOSY experiments. Once the formation of these cages was studied, Fujita and co-workers started to develop new derivatives based on the $M_{12}L_{24}$ framework as stellated cuboctahedron form in order to mimic the spherical virus capsids, the internal and external functionalization, and the capability to encapsulate targeted molecules as the virus capsid does.[80] More recently, this group reported the synthesis and studies of a self-assembled coordination cage as a molecular press, which can be used to flatten the bowl-shape guess corannulene with the inclusion of naphthalene diimide.[84] Another approach developed by these researchers is the conversion $M_{18}L_6$-$M_{24}L_8$ capsule-capsule induced by the inclusion of a guest, where a trigonal bipyramid is converted to an octahedron. The latter capsule possesses a cavity volume three times larger than the former.[85] In 2016 it was first reported a metal-ligand macrocycle constituted by $M_{30}L_{60}$ (icosidodecahedron) linked through Pd(II) ions and bipyridin thiophene ligands. The structure of this macromolecule was corroborated by X-ray crystallographic analysis and the interior space is large enough (157.000 $Å^3$) to confine proteins (Figure 17c).[86] In a subsequent publication in Nature, in the same year, they reported another $M_{30}L_{60}$ macromolecule self-assembled in solution through metal-ligand interactions, which consists in a combination of 8 triangles and 24 squares, and has the symmetry of a tetravalent Goldberg polyhedra.[87] The trivalent Goldberg polyhedra is naturally present in virus capsids and fullerenes, but this tetravalent form was not previously reported at the molecular level. Finally, the future target of this research group, is the obtaining of the largest self-assembled metal-ligand macrocycle, the rhombicosidodecahedron $M_{60}L_{120}$ (Figure 17d).

Multicomponent systems

Another interesting approach developed by Stang and co-workers was the formation of *multicomponent metallo-macrocycles* as it is depicted in Scheme 3a.[88] For such purpose, mixing carboxylate and pyridyl ligands with 90° Pt(II) in a proper ratio, coordination-driven self-assembly occurs, because Pt(II) prefers to bind one pyridine and one carboxylate ligand. This heteroleptic coordination allows for the selective generation of discrete multicomponent structures (2D and 3D) via charge separation on the metal centers. This new strategy can even induce the supramolecule-to-supramolecule transformations from a previously prepared homoleptic derivative to heteroleptic systems due to an external stimuli produced by the addition of the carboxylate ligand (Scheme 3b).

[84] B. M. Schmidt, T. Osuga, T. Sawada, M. Hoshino, M. Fujita, *Angew. Chem. Int. Ed.* **2016**, *55*, 1561–1564.
[85] S. Wang, T. Sawada, K. Ohara, K. Yamaguchi, M. Fujita, *Angew. Chem. Int. Ed*. **2016**, *55*, 2063–2066.
[86] D. Fujita, Y. Ueda, S. Sato, H. Yokoyama, N. Mizuno, T. Kumasaka, M. Fujita, *Chem*, **2016**, *1*, 91–101.
[87] D. Fujita, Y. Ueda, S. Sato, N. Mizuno, T. Kumasaka, M. Fujita, *Nature*, **2016**, *540*, 563–566
[88] a) Y. -R. Zheng, Z. Zhao, M. Wang, K. Ghosh, J. B. Pollock, T. R. Cook, P. J. Stang, *J. Am. Chem. Soc.* **2010**, *132*, 16873–16882; b) M. Wang, Y. -R. Zheng, T. R. Cook, P. J. Stang, *Inorg. Chem.* **2011**, 50, 6107–6113 c) J. B. Pollock, T. R. Cook, G. L. Schneider, P. J. Stang, *Chem. Asian J.* **2013**, *8*, 2423–2429.



Scheme 3. Self-assembled multicomponent systems developed by Stang. (**a**) Formation of a supramolecular cage and (**b**) self-assembly of a trigonal prism *via* supramolecule-to-supramolecule transformation.

3.2 Macrocycles Assembled *via* H-Bonding.

H-bonding is one of the most popular driving force to form macrocycles among other non-covalent interactions,[73,75c,89] such as metal-ligand interactions, since it allows a high degree of control over supramolecular architectures due to its high selectivity and directionality.[34a,90] Depending on the binding strength of their motifs, H-bonded assemblies are stable only at a well-defined range of monomer concentrations, temperature and solvent environments.

In this section we will first expose a few selected examples collected from the literature where a given molecule with two binding sites featuring different H-bonding arrays form well-defined macrocycles in solution.[91] Sijbesma, Meijer and co-workers reported one of the examples to form macrocycles in solution constituted by two ditopic monomers taking the advantage of the strong *DDAA-AADD* quadrupole H-bond motifs in the UPys (**1**, Figure 18a).[92] Zimmerman and Duerr reported in 1992 the cyclotrimerization of pyrido[4,3g]-quinolinedione (**2**, Figure 18b).[93] This self-assembled trimer is formed by the cooperative interaction of six H-bonds. The high stability of this assembly was corroborated by VPO and concentration dependent ^{1}H NMR measurements in chloroform. Ballester and Mendoza[94] recently analyzed the *EM* value of the system as 740 M, which meets the conditions of quantitative assembly

[89] F. Biedermann, W. M. Nau, H.-J. Schneider, *Angew. Chem. Int. Ed.* **2014**, *53*, 11158–11171.
[90] a) D. González-Rodríguez, A. P. H. J. Schenning, *Chem. Mater.* **2011**, *23*, 310–325; b) A. P. H. J. Schenning, D. González-Rodríguez, *Organic Nanomaterials: Synthesis, Characterization, and Device Applications* (Eds.: T. Torres, G. Bottari), John Wiley & Sons, Inc., Hoboken, New Jersey, **2015**, 33–58; c) T. Steiner, *Angew. Chem. Int. Ed.* **2002**, *41*, 48–76; d) *Hydrogen Bonded Supramolecular Structures* (Eds.: Z.-T. Li, L.-Z. Wu), Springer-Verlag, Berlin Heidelberg, **2015**.
[91] M. J. Mayoral, N. Bilbao, D. González-Rodríguez, *ChemistryOpen*, **2016**, *5*, 10–32.
[92] A. Tessa ten Cate, H. Kooijman, A. L. Spek, R. P. Sijbesma, E. W. Meijer, *J. Am. Chem. Soc.* **2004**, *126*, 3801–3808.
[93] S. C. Zimmerman, B. F. Duerr, *J.Org. Chem.* **1992**, *57*, 2215–2217.
[94] P. Ballester, J. de Mendoza in Modern Supramolecular Chemistry (Eds.: F. Diederich, P. J. Stang, R. R. Tykwinski, Wiley-VCH, Wein-heim, **2008**, 69–111.

defined by Ercolani (K_{ref}·*EM* > 185·n).[73c] Sleiman and colleagues published a new strategy to control the formation of cyclic tetramers in solution by using an external stimuli (**3**, Figure 18c).[95] For such propose, the central block used was a photoresponsive azobenzene unit, which can undergo a reversible *trans-cis* photoisomerization. The self-recognition moieties are constituted by carboxylic acids, which forms linear H-bonding interactions. In the photogenerated *cis*-form, the carboxylic acids are oriented to perpendicular fashion, leading to a closed system structure. In contrast, in the *trans*-isomer, due to the alignment of both, recognition units and central block, the monomer self-assemble into linear aggregates. These self-assembly behavior could be corroborated by transmission electron microscopy (TEM) and dynamic light scattering (DLS). Also, these techniques revealed a second level of organization suggesting a further organization of the supramolecular cycles through π-π staking and/or alkyl-alkyl interactions.

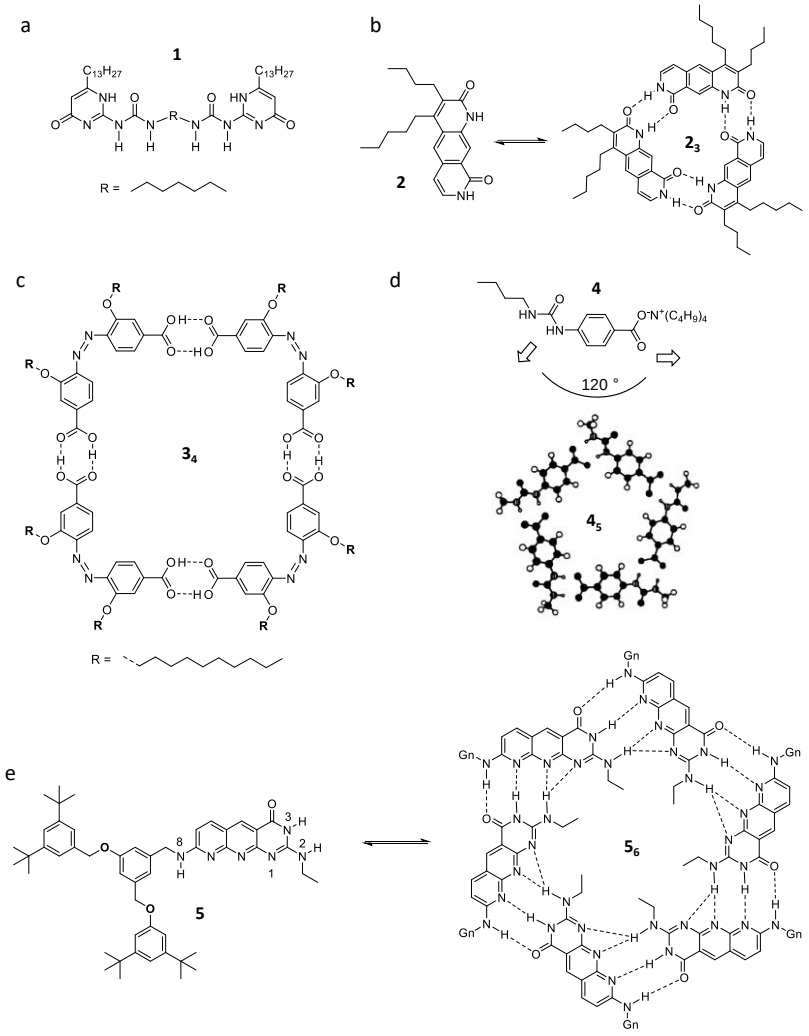


Figure 18. H-bonded macrocycles composed by a) two members, b) three members, c) four members, d) five members and e) six members.

[95] F. Rakotondradany, M. A. Whitehead, A.-M. Lebuis, H. F. Sleiman, *Chem. Eur.J.* **2003**, *9*, 4771–4780.

Hamilton and his co-workers reported one of the few examples of five-membered hydrogen-bonded macrocycles presented in the literature (Figure 18d).[96] In this work the researchers studied the capability of molecule **4** to form self-assembled structures depending on the conditions and the nature of the interacting components. Due to the exact angle of 120° between recognitions sites featured in this molecule, both cyclic hexamers and pentamers can be formed from the planar and nonplanar disposition of the H-bonding groups, respectively. However, the data analyzed form concentration-dependent ^{1}H NMR experiments in different solvents over a concentration range 0.1-100 mM fitted best with a monomer-pentamer equilibria. As the last example, using a bifunctional monomer bearing a complementary *DDA-AAD* H-bonding motif, Kolotuchin and Zimmerman designed a more robust hexameric aggregate (Figure 18e).[97] These discrete macrocycle is stablished by 18 H-bonds and additionally supported for six secondary H-bonds involving the 2-NH groups. In order to assess the robustness of this complex, solvent-dependant ^{1}H NMR experiments were carried out. These experiments shows that the cyclic assembly **5** can survive in a 15% aqueous THF. In related publications,[98] the researchers changed the substituent in the N-8 by two different generation of dendrimers where the stability of the hexameric aggregates decreases upon addition of DMSO-d_6 in consequence of the steric effects between the peripheral bulky groups.

One of the first ***molecular boxes*** linked only by H-bonds was reported by Rebek and his colleagues in 1993.[99] In this case two mutually complementary curved building blocks allow to prepare hollow capsules in solution (Figure 19a). Latter, Reinhoudt and co-workers[100] developed a macromolecule constituted by three units of their calix[4]arene derivative (**6**-**8**) with six equivalents of the merocyanine dye and stabilised by the highly cooperative action of 36 H-bonds (Figure 19b). Finally, Ballester and his colleagues more recently have been developed an extensively research of supramolecular boxes based in calix[4]pyrroles and resorcine[4]arenes among others.[101] A representative example is the aromatic calix[4]pyrroles cavities functionalized with four ureas in the *para* position of their *meso* phenyl substituents.[101b] The tetraurea calix[4]pyrrole dimerizes reversibly forming a cyclic array of 16 hydrogen bonds in dichloromethane solution in the presence of one molecule of 4,4'-bipyridine *N*-*N*'-dioxide and encapsulating one molecule of bis-*N*-oxide. The encapsulated guest is bound in the cavity by hydrogen bonding to the two endohedral calix[4]pyrrole centers (Figure 19c).

[96] A. Zafar, S. J. Geib, Y. Hamuro, A. D. Hamilton, *New J. Chem*. **1998**, *22*, 137–141.
[97] S. V. Kolotuchin, S. C. Zimmerman, *J. Am. Chem. Soc.* **1998**, *120*, 9092–9093.
[98] a) S. C. Zimmerman, F. Zeng, D. E. C. Reichert, S. V. Kolotuchin, *Science*, **1996**, *271*, 1095–11098; b) P. S. Corbin, L. J. Lawless, Z. Li, Y. Ma, M. J. Witmer, S. C. Zimmerman, *Proc. Natl. Acad. Sci. U.S.A*. **2002**, *99*, 5099–5104; c) Y. Ma, S. V. Kolotuchin, S. C. Zimmerman, *J. Am. Chem. Soc.* **2002**, *124*, 13757–13769.
[99] R. Wyler, J. de Mendoza, J. Rebek, *Angew. Chem. Int. Ed.* **1993**, *32*, 1699–1701.
[100] L. J. Prins, C. Thalacker, F. Wurthner, P. Timmerman, D. N. Reinhoudt, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10042–10045.
[101] a) G. Gil-Ramíez, J. Benet-Buchholz, E. C. Escudero-Adán, P. Ballester, *J. Am. Chem. Soc.* **2007**, *129*, 3820–3821; b) P. Ballester, G. Gil-Ramírez, *Proc. Natl. Acad. Sci. U.S.A*. **2009**, *106*, 1045–10459; c) L. Osorio-Planes, M. Espelt, M. A. Pericàs, P. Ballester, *Chem. Sci.* **2014**, *5,* 4260–4264; d) A. Galán, V. Valderrey, P. Ballester, *Chem. Sci.* **2015**, *6*, 6325–6333 e) A. Galán, G. Aragay, P. Ballester, *Chem. Sci.* **2016**, *7,* 5976–5982.

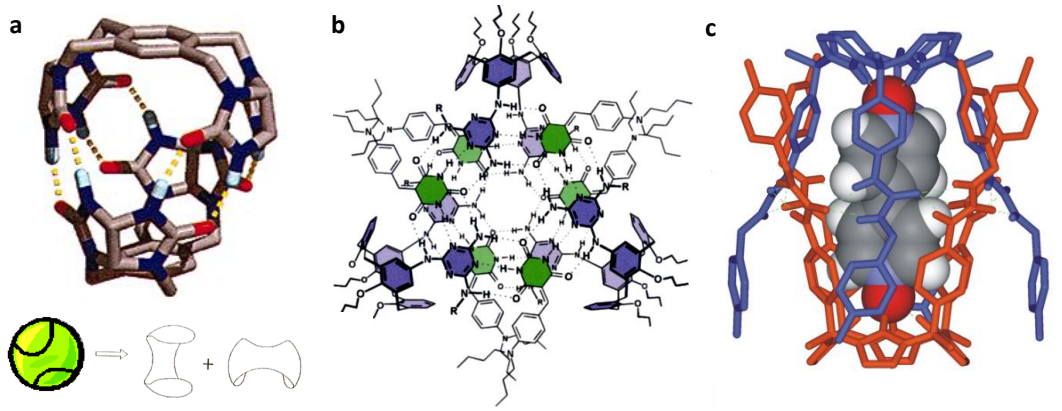


Figure 19. Schematic representation and molecular structure of supramolecular H-bonded cages developed by (**a**) Rebek, (**b**) Reinhoudt and (**c**) Ballester.

As it has been commented previously, nucleobases are one the most popular motifs to develop H-bonded complexes and discrete assemblies. Nucleobases are an interesting and powerful tool due to their versatility, and researchers have been able to synthesize different structures such as macrocycles constituted from two to six members, nucleobase carriers, logic gates, photoinduced energy transfer systems, nucleobase-bonded oligopeptides, supramolecular polymers and soft materials.[102] Being stronger and unsymmetric, the *ADD-DAA* H-bonding interaction between G and C is preferable to the *DA-AD* one between A and U or the *DAD-ADA* one between 2-aminoadenosine and U.

In the study of the formation in solution of discrete ***cyclic dimers***, Sessler and his co-workers designed and synthesized different monomers containing two complementary nucleobases with different spacing blocks in order to study their stability in polar solvents. In order to understand the influence of the linker geometry and rigidity on the association strength and the dimerization fidelity, a series of G-C and A-U ditopic monomers were synthesized.[103] ^{1}H NMR titrations of the respective mononucleobase derivatives, as well as dilution experiments of dinucleobase compound**s** (Figure 20) in DMSO-d_6 did not reveal a significant increase in the dimerization constant. This unexpected behavior could be due to both the high number of degrees of freedom in the flexible system (**9**, Figure 20a), which must imply a large entropic cost upon cyclization and hence a low *EM* value.[103a] Based on these results, the authors synthesized more rigid systems by using 1,8-diethynylanthracene as the spacer and complementary A-U base pairing entities as the recognition units (**10**, Figure 20b).[100b] The rigid homodimers of AU dinucleotide have an improved stability compared to the monomeric base pairs due to the preorganization of the system, whereas the use of flexible spacers hardly improved dimer stability. When bulkier substituents are placed in the same AU homodimer, the steric hindrance leads the system to dissociate with a minor amount of DMSO-d_6 in $CDCl_3$ than the first one. The resulting homodimer when the central block was changed by a dibenzofuran moiety between A and U nucleobases was even less stable due to a less optimized preorganization.[103c] The last generation of these derivatives concern to a

[102] M. J. Mayoral, C. Montoro-García, D. González-Rodríguez, *Self-assembled Systems via Nucleobase Pairing,* Book chapter in *Comprehensive Supramolecular Chemistry II*, Elsevier Ltd., UK. **2017**, DOI:10.1016/B978-0-12-409547-2.12536-3.
[103] a) J. L. Sessler, D. Magda, H. Furuta, *J. Org. Chem.* **1992**, *57*, 818–826; b) J. L. Sessler, R. Wang, *J. Am. Chem. Soc.* **1996**, *118*, 9808–9809; c) J. L. Sessler, R. Wang, *Angew. Chem. Int. Ed.* **1998**, *37*, 1726–1729; d) J. L. Sessler, R. Wang, *J. Org. Chem.* **1998**, *63*, 4079–4091.

stronger G-C homodimer with bulky substituents, which surprisingly was even less stable than the AU ditopic monomer.[103d] All of these features indicated that either the lower structural preorganization, or the presence of bulky substituents, which cause severe steric hindrance in the dimers, leads to a significant reduction in the dimer stability. It is interesting to note, when comparing the four systems, that the kinetic behavior is also strongly affected by their different geometric features and steric hindrances. While the more stable dimer $\mathbf{10}_2$ exhibited a slow equilibrium with the monomer in the NMR timescale in $CDCl_3$-DMSO-d_6 mixtures, the rest of the assemblies were in the fast exchange regime at room temperature.

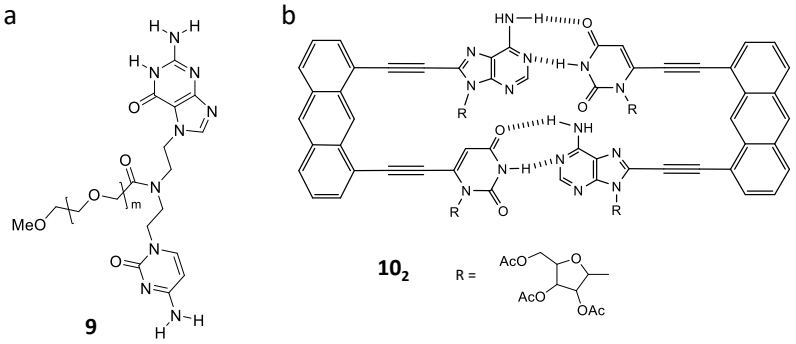


Figure 20. Dimeric structures based on nucleobase derivatives

Other dimeric systems assembled from dinucleobase precursors have been studied by other researchers. For instance, assemblies formed by two crown ethers disubstituted with A and/or T at both ends, and thus resembling molecular boxes, were studied by Gokel and colleagues (Figure 21a,b).[104] In order to confirm the proposed molecular box assembly mode, ^{1}H NMR studies as a function of both concentration and temperature, as well as NOE and VPO measurements, were performed. In $CDCl_3$, ^{1}H NMR evidence for the formation of A-T H-bonds was taken from the downfield shift of the T N–H imide signal in the presence of the A counterpart, and finally corroborated by the detection of NOE cross-peaks signals that suggested a mixture of Watson-Crick and Hoogsteen H-bonding interactions (Figure 21a,b). It is interesting to note that these bonded conformations could be further stabilized, through a templation effect, by intramolecular bond formation with alkyldiamine salts *via* complexation of the crown moieties. This strong association was also detected by ESI-MS spectra.[105]

[104] a) M. Kim, G. W. Gokel, *J. Chem. Soc. Chem. Commun.* **1987** 1686–1688; b) O. F. Schall; G. W. Gokel, *J. Am. Chem. Soc.* **1994**, *116*, 6089–6100.
[105] K. Wang, O. F. Schall, G. W. Gokel, *Supramol. Chem.* **1996**, *7*, 85–90.

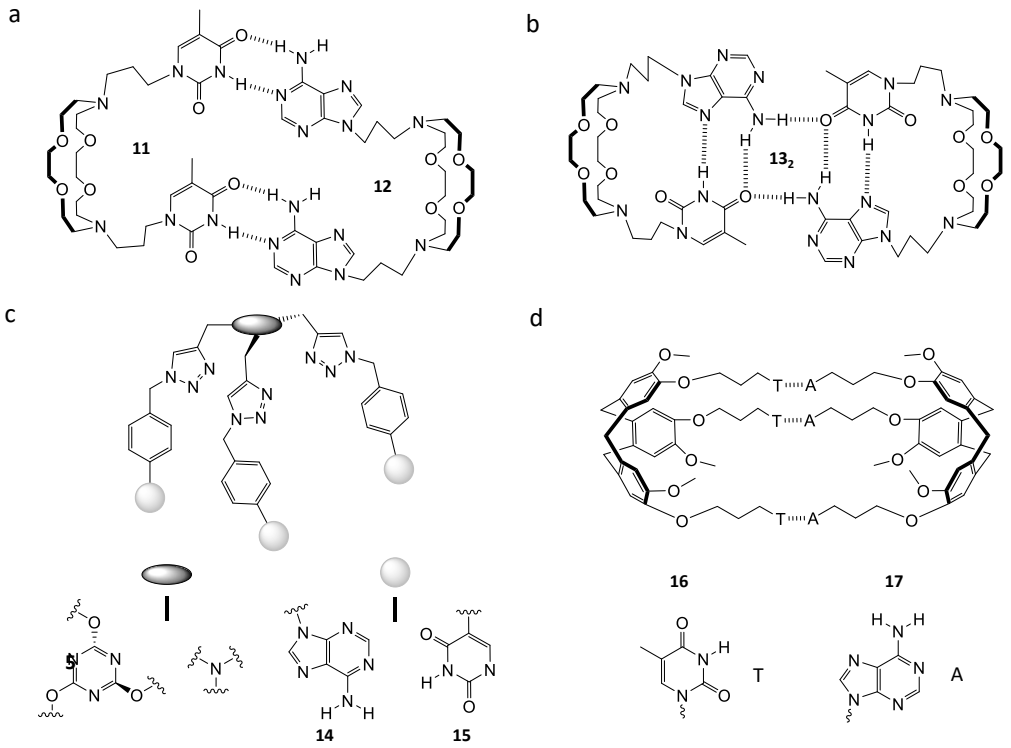


Figure 21. Supramolecular cages based on Watson-Crick H-bonding interactions developed by a,b) Gokel,[104,105] c) Grosu[106] and d) Purohit.[107]

A different kind of molecular box was developed by Grosu and co-workers that makes use of tripodands, featuring A and T nucleobases, as building blocks (Figure 21c).[106] Unfortunately these compounds are soluble only in DMSO, which competes strongly for H-bonding, whereas the addition of little amounts of non-polar solvents led to precipitation. Due to this insolubility problem, X-ray analysis could not be carried out and the only way to confirm the association was limited to electrospray ionization high resolution mass spectrometry (ESI-HRMS), using the positive ionization mode, where the formation of self-associated dimers and trimers was confirmed. Related work was recently carried out by Purohit and co-workers with the aim of forming dimeric capsules with cyclotriguaiacyclene derivatives tethered to A and T (**16**·**17**; Figure 21c).[107] When the blocks were mixed in a 1:1 ratio, the resulting supramolecular box formation was complete, as corroborated by HR-MS, diffusion-ordered spectroscopy DOSY, and ^{1}H-NMR studies. An interesting feature to note is that the solubility of the 1:1 **16**·**17** mixture was enhanced with respect to the individual compounds, in contrast to what was observed with molecules **14** and **15**.

As a last example, Vasella and co-workers developed another approach to synthesize a large family of A-U, U-A, C-G and G-C dinucleosides that they named as “Oligonucleotide Analogues with Integrated Bases and Backbones (ONIBs)”. In these studies, the nature of the linking unit was modified in order to understand its influence in duplex formation, which was

[106] M. L. Golban, V. Paşcanu, N. D. Hădade, L.Pop, C. Socaci, I. Grosu, *Synthesis*, **2014**, *46*, 1229–1235.
[107] P. Satha, G. Illa, A. Ghosh, C. S. Purohit, *RSC Adv.* **2015**, *5*, 74457.

thoroughly studied by a wide number of techniques, namely NMR, ESI-MS, Vapor Pressure Osmometry (VPO), Shift/Concentration Curves (SCC) and CD spectroscopy.[108]

There is only one example of a ***three-membered*** macrocycle H-bonded through Watson-Crick interactions, developed by Sessler and colleagues.[109] Following the previous strategy of designing artificial guanosine-cytidine dinucleosides to form cohesive dimers, a novel planar derivative (**18**) was synthesized by adding a double bond between G and C nucleobases, so as to extend the supramolecular cyclization to larger cycles (Figure 22). ^{1}H-NMR studies in $CDCl_3$ evidenced the self-assembly process. The guanosine amide proton (NH-1) was found at 13.82 ppm, which is characteristic of strong H-bonding, and this shift was not affected by concentration-dependent measurements at room temperature. Nuclear Overhauser Effect Spectroscopy (NOESY) spectra in $CDCl_3$ showed strong cross-peaks between the guanosine amide proton and the cytidine amino proton (NH-4) indicating a close spatial arrangement and thus being consistent with Watson-Crick base-pairing between individual **18** monomers. Besides, ESI-MS and size-exclusion chromatography (SEC) experiments (which exhibited only one peak in THF) suggested that the mass and size of the resulting supramolecular species was consistent with a trimeric structure (**18**$_3$).

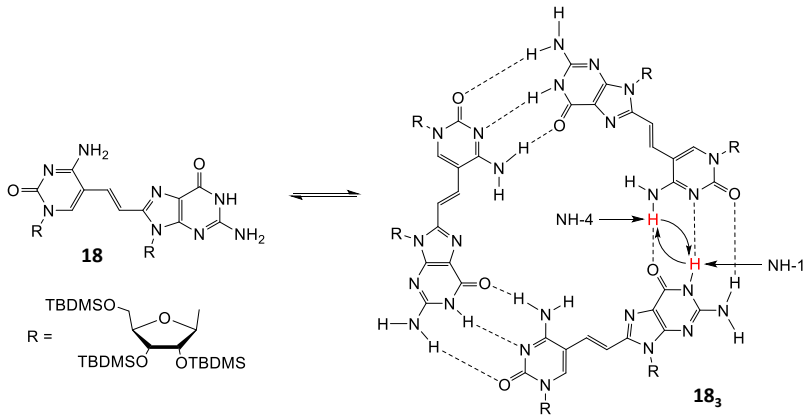


Figure 22. Monomer **18** and its supramolecular H-bonded trimer designed by Sessler *et al.*[109]

Another remarkable approach to form stable macrocycles is represented by the ***Janus-Type molecules***, where two nucleobase motifs (commonly the G-C base pair) are fused in the same heterobicycle. Therefore, two well defined Watson-Crick sides are defined within the same core in this kind of derivatives, each of them having different functional groups and recognition H-bonding patterns. The groups of Lehn, Mascal, Fenniri, Perrin and He have made important synthetic contributions to these kinds of Janus-type molecules and different supramolecular aspects have been investigated.

Perrin *et al.* synthesized the first example of a self-assembled ***tetrameric rosette*** containing the G-C pattern (**19**; Figure 23)[110], in which a central pyrrole heterocycle orients the H-bonding faces of both G (ADD) and C (DAA) at a 90° angle. Watson-Crick association

[108] a) K. Chiesa, B. Bernet, A. Vasella, *Helv. Chim. Acta,* **2010**, *93*, 1822–1843; b) X. Zhang, B. Bernet, A. Vasella, *Helv. Chim. Acta,* **2007**, *90*, 891–908; c) M. Schulze-Adams, D. Touboul, B. Bernet, A. Vasella, *Helv. Chim. Acta*, **2014**, *97*, 1037–1054, and references cited therein; d) B. Bernet, Z. Johar, A. Ritter, B. Jaun, A. Vasella, *Helv. Chim. Acta*, **2009**, *92*, 2596–2630; e) N. Bogliotti, A. Vasella, *Helv. Chim. Acta*, **2010**, *93*, 888–909.
[109] J. L. Sessler, J. Jayawickramarajah, M. Sathiosatham, C. L. Sherman, S. Brodbelt, *Org. Lett.* **2003**, *5*, 2627–2630.
[110] A. Asadi, B. O. Patrick, D. M. Perrin, *J. Am. Chem. Soc.* **2008**, *130*, 12860–12861.

then leads to a stable tetrameric cycle associated by 12 H-bonds (Figure 23a). Temperature-dependent ^{1}H-NMR measurements from 25 to -70 °C of a solution of **18** in DMSO-d_6/$CDCl_3$ corroborated H-bonding between the two faces of the monomer (Figure 23b). At higher temperatures, the 2-NH_2 group of the G-face rotates rapidly on the NMR time scale, and both protons appeared as a single broad coalesced resonance at 6.3 ppm. At low temperatures (-65 °C) they were resolved as two distinct peaks at 5.75 and 7.3 ppm. In contrast, the amino protons of the 4-NH_2 C-face appeared as a very broad, almost undetectable resonance signal between 6.75 and 7.65 ppm which, at -10 °C, rapidly splits into two well-resolved peaks at approximately 6.8 and 7.5 ppm. At -65 °C the protons in the two amino groups present thus four distinct resonances. The observation of new peaks at very low temperatures is consistent with a G-C pairing scheme and suggested the formation of a tetrameric rosette in solution. DOSY experiments as well as ESI-MS analysis of **19** confirmed such association. Neither a peak for the corresponding trimer nor for any other higher-order aggregate was detected.

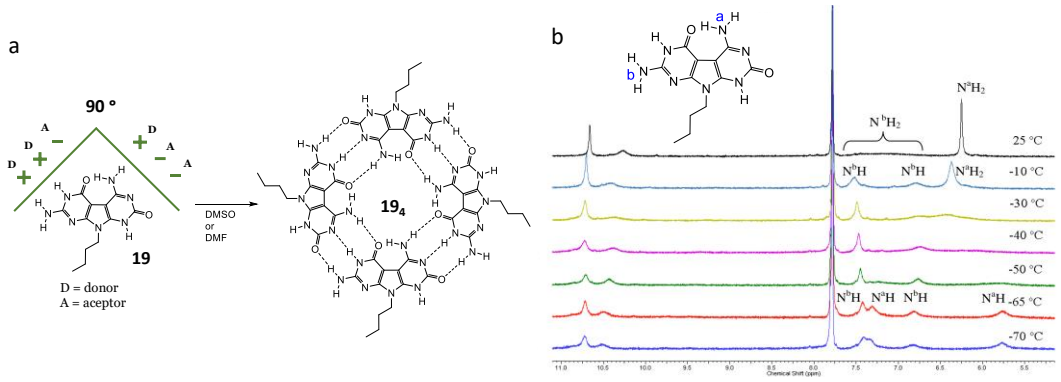


Figure 23. a) Cyclic tetramer formation from a Janus-type G-C monomer (**19**) designed by Perrin and co-workers. b) Region of the ^{1}H NMR spectra of **19** at different temperatures.

In 2011, Butkus, Wärnmark and colleagues reported the first selective assembly by tautoleptic aggregation of an enantiomerically pure cavity that is a supramolecular belt, from one enantiomerically pure GC monomer containing one inherently non-self-complementary motif.[111]

The groups of Lehn[112] and Mascal[113] pioneered the synthesis of Janus-type molecules like **20**, **21** and **22** by fusing two ***six-member heterocycles*** which contain the H-bonding codes of both C (ADD) and G (DAA) oriented with a 60 ° angle. This molecular geometry mediates their self-organization into hexameric supramolecular macrocyclic structures or rosettes (Figure 24). The quantitative estimation of the association constant of **20** and **21** was hampered by solubility problems. In $CHCl_3$, ^{1}H-NMR measurements showed a downfield shift of the NH protons and VPO measurements with a 2600 ± 10% molecular mass, confirmed that the desired cyclic hexamer is formed in solution.[102] Moreover, X-ray crystallographic studies gave proof for the presence of the cyclic hexamer $\mathbf{22}_6$ in the solid state (Figure 24c).[113a,b] In

[111] E. Orentas, C. -J. Wallentin, K. -E. Bergquist, M. Lund, E. Butkus, K. Wärnmark, *Angew. Chem. Int. Ed*. **2011**, *50*, 2071–2074.
[112] A. Marsh, M. Silvestri, J. -M. Lehn, *Chem. Commun*. **1996**, 1527–1528.
[113] a) M. Mascal, N. M. Hext, R. Warmuth, M. H. Moore, J. P. Turkenburg, *Angew. Chem. Int. Ed.* **1996**, *35*, 2204–2206; b) M. Mascal, N. M. Hext, R. Warmuth, J. R. Arnall-Culliford, M. H. Moore, J. P. Turkenburg, *J. Org. Chem.* **1999**, *64*, 8479–8484; c) M. Mascal, S. C. Farmer, R. Arnall-Culliford, *J. Org. Chem.* **2006**, *71*, 8146–8150.

this crystal structure, $\mathbf{22}_6$ hexamers overlapped each other to describe a highly porous solid with interwoven network channels (Figure 24d). This work represents a magnificent example where suitable monocrystals for X-ray analysis have been obtained in which the resolved structure faithfully represents the cyclic assembly found in solution. The rigid nature and full preorganization of the molecule towards cyclization avoid the formation of open oligomers and other H-bonded species driving the system to the formation of a well-ordered networks at the solid state.

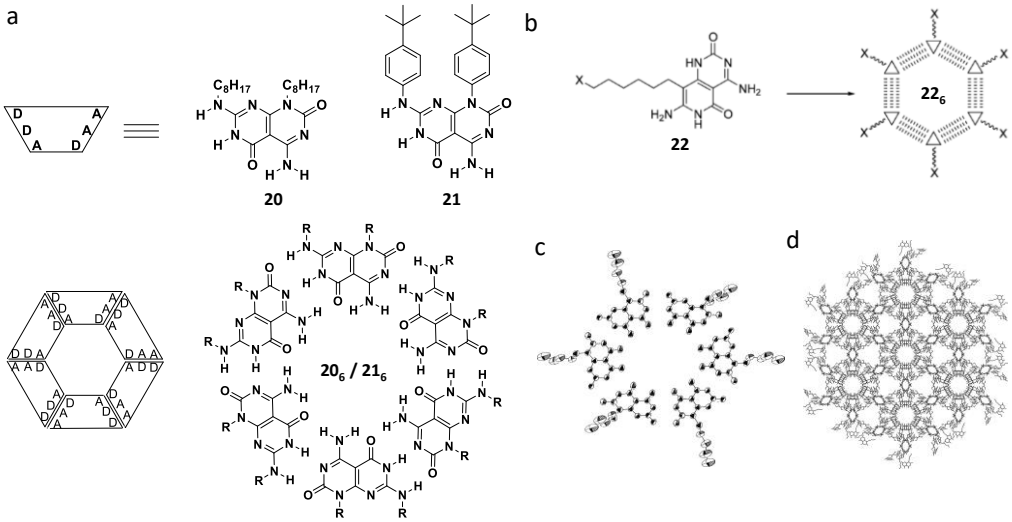


Figure 24. (a,b) Janus-type six membered rosettes $\mathbf{20}_6$, $\mathbf{21}_6$ and $\mathbf{22}_6$ synthesized by Lehn[112] and Mascal.[113] (c,d) Organization of compound **22** in the crystal.

This G-C Janus-type motif that associates into six-membered rosettes was later exploited by the Fenniri group to construct self-assembled ***rosette nanotubes***. This research group created novel heterobicyclic G-C molecules, like **23** (Figure 25), substituted by pendant groups that provided solubility either in water or in organic solvents. Because of the asymmetry of the Watson-Crick DAA-ADD H-bonding edges, the G-C monomer faithfully undergoes self-assembly in water[114] as well as in organic solvents[115] forming six-membered macrocycles joined by 18 H-bonds. Then, these hexameric rosettes produce hierarchical tubular stacks (rosette nanotubes, RNTs) where dimensions and properties can be modulated, as demonstrated by the authors in manifold AFM, TEM, and SEM microscopy studies.

[114] a) H. Fenniri,; P. Mathivanan, K. L. Vidale, D. M. Sherman, K. Hallenga, K. V. Wood, J. G. Stowell, *J. Am. Chem. Soc.* **2001**, *123*, 3854–3855; b) H. Fenniri, B. -L. Deng, A. E. Ribbe, K. Hallenga, J. Jacob, P. Thiyagarajan, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6487–6492; c) J. G. Moralez, J. Raez, T. Yamazaki, R. K. Motkuri, A. Kovalenko, H. Fenniri, *J. Am. Chem. Soc.* **2005**, *127*, 8307–8309; d) H. Fenniri, B. -L. Deng, A. E. Ribbe, *J. Am. Chem. Soc.* **2002**, *124*, 11064–11072; e) B. -L. Deng, R. L. Beingessner, R. S. Johnson, N. K. Girdhar, C. Danumah, T. Yamazaki, H. Fenniri, *Macromolecules*, **2012**, *45*, 7157−7162.

[115] a) G. Tikhomirov, M. Oderinde, D. Makeiff, A. Mansouri, W. Lu, F. Heirtzler, D. Y. Kwok, H. Fenniri, *J. Org. Chem.* **2008**, *73*, 4248–4251; b) G. Tikhomirov, T. Yamazaki, A. Kovalenko, H. Fenniri, *Langmuir*, **2008**, *24*, 4447–4450; c) A. Durmus, G. Gunbas, S. C. Farmer, M. M. Olmstead, M. Mascal, B. Legese, J. -Y. Cho, R. L. Beingessner, T. Yamazaki, H. Fenniri, *J. Org. Chem*. **2013**, *78*, 11421–11426.

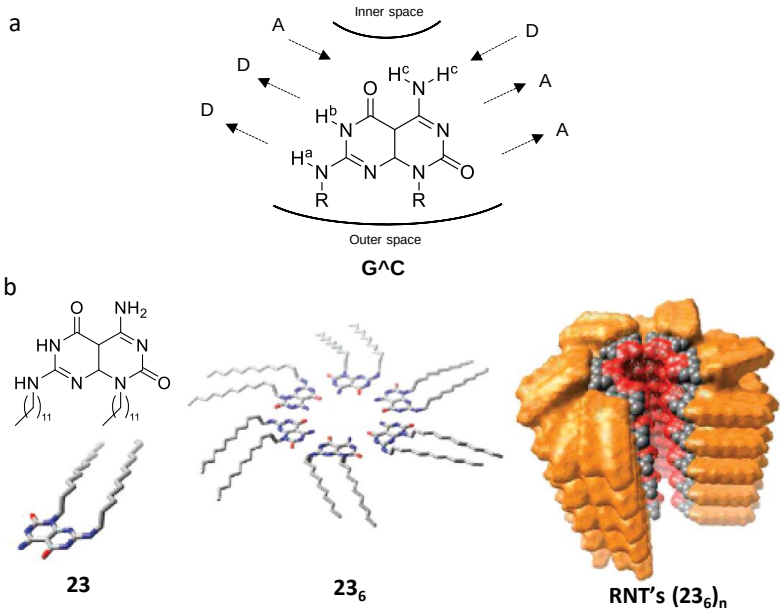


Figure 25. a) General structure of a G-C Janus-type monomer **23** that b) self-assembles in solution into hexameric H-bonded macrocycles that then stack leading to rosette nanotubes (RNTs).

The peripheral diameter and properties of these RNTs can be dictated by the choice of the functional groups conjugated to the Janus-type motif, whereas the inner space is directly related to the distance separating the G-C H-bonding arrays within the monomer.[116] For instance, the authors synthesized a new self-complementary monomer with a larger cavity adding a pyridine ring between the G- and C-like interfaces (Figure 26a). Compared to **24**, the self-assembling rosette nanotube **25** showed a higher diameter and a larger π system, which was proposed to allow electronic transport along the main axis (Figure 26b). Images of randomly oriented RNTs displayed an average outer diameter of 3.5 ± 0.3 nm. Thus, the inner and outer diameters of **25** RNTs increased by 0.4 nm in comparison to relative to **24** RNTs (Figure 26c-d).[114e]

[116] a) G. Borzsonyi,; R. S. Johnson, A. J. Myles, J. -Y. Cho, T. Yamazaki, R. L. Beingessner, A. Kovalenko, H. Fenniri, *Chem. Commun*. **2010**, *46*, 6527–6529; b) G. Borzsonyi, R. L. Beingessner, T. Yamazaki, J. -Y. Cho, A. J. Myles, M. Malac, R. Egerton, M. Kawasaki, K. Ishizuka, A. Kovalenko, H. Fenniri, *J. Am. Chem. Soc*. **2010**, *132*, 15136–15139.

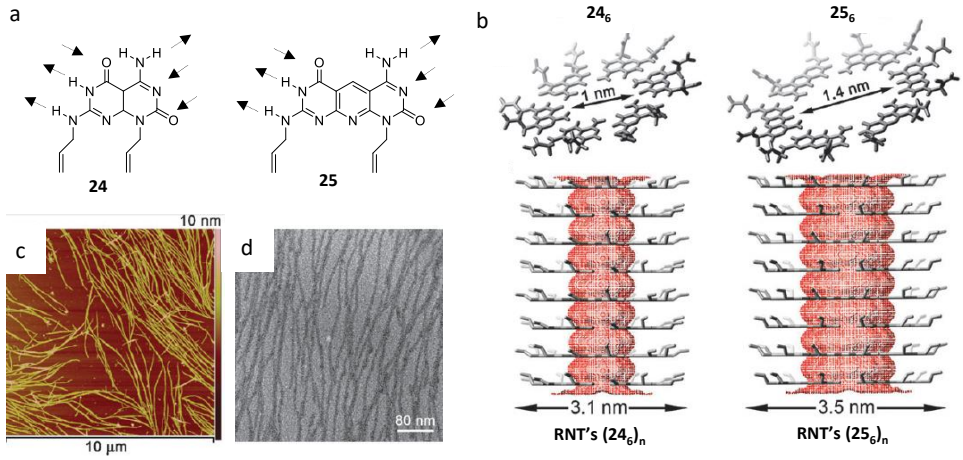


Figure 26. a) G-C fused monomers **24** and **25** and b) their RNTs models. Selected regions of c) TM-AFM and d) SEM microscopy images of self-assembled **25**.

Some of the latest developments of the group focused on the design of RNTs with chirooptical[117]and electron-donor behavior,[118] as well as on the covalent fixation of the self-assembled nanotubes.[114e] In the first case they relied on the ability of a single chiral molecule to express multiple supramolecular chirality outputs as a result of preferred conformational states under a set of physical stimuli. These so-called "chiromers" are supramolecular conformational isomers that a) are thermodynamically stable, b) can memorize their chirality and c) can amplify their chirality in an achiral environment. In 2015, these researchers synthesized new RNTs substituted with porphyrins or oligothiophene units, behaving as electron-donor materials in combination with $PC_{61}BM$ fullerene as an electron-accepting material.[118] Finally, the functionalization of the heterobicyclic unit with alkyldiamine bifunctional units allowed the covalent capture of the RNTs in the presence of adipoyl chloride, resulting in films composed of RNT fibers.[114e]

[117] a) R. S. Johnson, T. Yamazaki, A. Kovalenko, H. Fenniri, *J. Am. Chem. Soc.* **2007**, *129*, 5735–5743; b) U. D. Hemraz, M. El-Bakkari, T. Yamazaki, J. -Y. Cho, R. L. Beingessner, H. Fenniri, *Nanoscale*, **2014**, *6*, 9421–9427.
[118] L. Shuai, V. Parthasarathy, J. -Y. Cho, T. Yamazaki, R. L. Beingessner, H. Fenniri, *MRS Proceedings*, **2015**, *1737*, mrsf14-1737-u01-04.

Background and Objectives

Background in *MSMn.*

1) Nanostructured Molecular Systems and Materials (MSMn).

This Thesis has been carried out in the *Nanostructured Molecular Systems and Materials* (*MSMn*) group at the Universidad Autónoma de Madrid. The research in *MSMn* aims at a common final objective: to improve or create new functions in organic materials by rationally ordering functional molecules at the nanoscale using the tools of self-assembly, or in other words, using noncovalent synthesis. One of the main projects in the group, *Programmed Nanostructuration of Organic Materials*, employs the structure of a self-assembled cyclic tetramer as a common feature in different subprojects so far (Figure 27): (1) the study of non-covalent macrocyclizations; (2) the development of π-conjugated functional self-assembled systems; (3) the formation of nanostructured porous surfaces and networks; (4) the polymerization of these cyclic tetramers into nanotubes with custom-tailored pores; and (5) the transfer of this supramolecular motif to biopolymers, to create tubular quadruplex DNA.

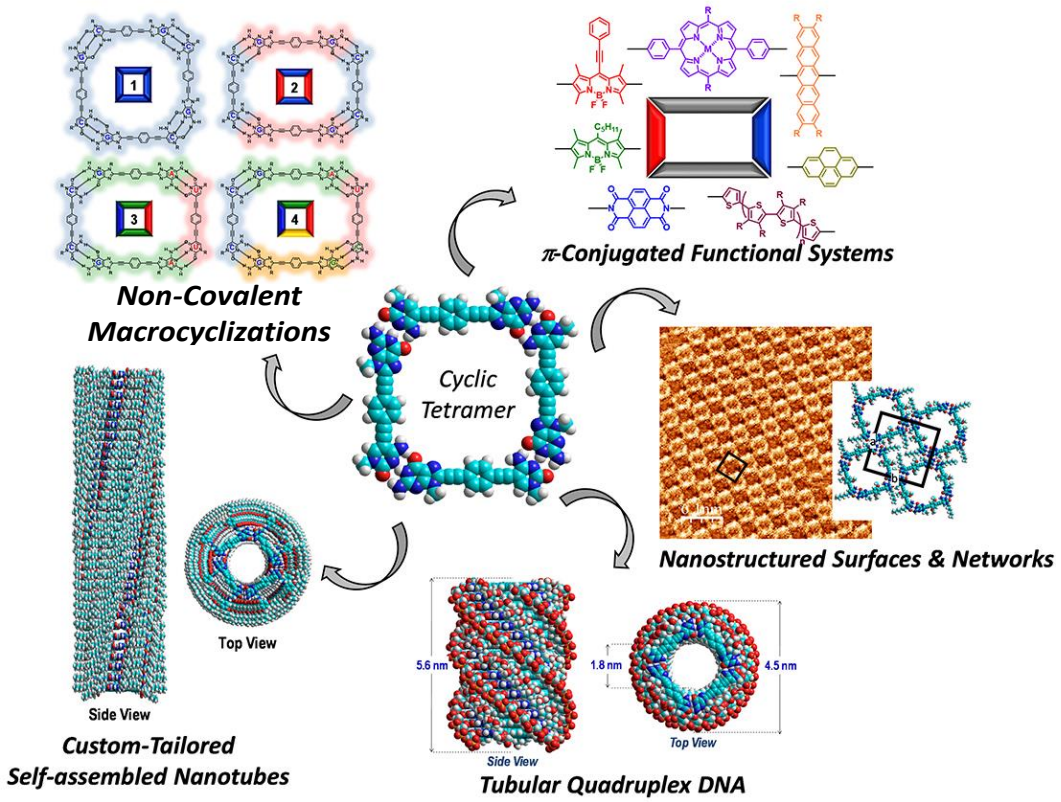


Figure 27. Programmed Nanostructuration of Organic Materials.

2) The Cyclic Tetramer.

This common macrocycle is built *via* H-bonding interactions between four monomer molecules (Figure 28). The monomers basically comprise a *rigid π-conjugated* ***central block*** that is linearly disubstituted at both ends through suitable ***spacers*** with *self-assembling* ***directors,*** able to interact by H-bonding.

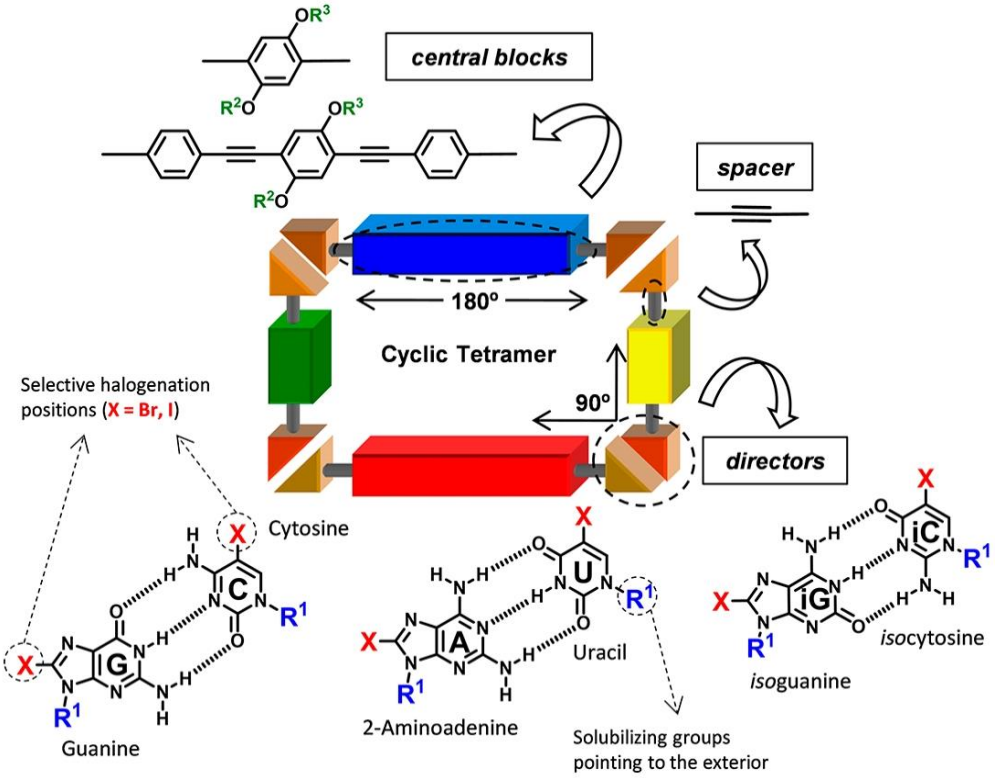


Figure 28. Key structural elements that lead to cyclic tetramer assembly.

The H-bonding ***directors*** are actually nucleobases, *i.e.* naturally occurring guanine (G), cytosine (C) and uracil (U); and non-natural 2-aminoadenine (A),[119] *iso*guanine (iG) and *iso*cytosine (iC). Together, they constitute a family of three complementary Watson-Crick couples: G–C, A–U, and iG–iC, that associate *via* triple H-bonding patterns. That is the reason why adenine has been replaced by 2-aminoadenine, so as to keep the three-fold H-bonding pattern and stabilize self-assembled structures. When these bases are subjected to electrophilic halogenation reactions, purines are substituted at the 8-position and pyrimidines at the 5-position (marked as **X** in Figure 28). It is essential to note that when the purine-pyrimidine pairs interact *via* complementary Watson-Crick H-bonding, these two positions form an exact 90° angle. On the other hand, the marked **R^1** positions, the 9-position in purines and the 1-position in pyrimidines, are always pointing toward the exterior of the nucleobase pairs. These are the same positions where the riboses are placed in RNA, and they can be further utilized to incorporate key solubilizing groups in order to approach different purposes for the cyclic tetramers.

[119] 2-Aminoadenine (or 2,6-diaminopurine) will be hereafter abbreviated as A, for the sake of simplicity.

The ***central blocks*** have to be rigid units substituted at both ends with an exact angle of 180º. They will typically comprise π-conjugated units that carry reactive groups at opposite positions and that can be endowed with diverse functions and equipped with different lateral groups (**R²** and **R³** in Figure 29), depending on the application requirements. There is a large number of functional π-conjugated units that fulfil the prerequisite of linear di-substitution: TTF, conjugated oligomers, aromatic acenes, perylenes, ABAB porphyrin or phthalocyanine macrocycles, etc. (Figure 29).

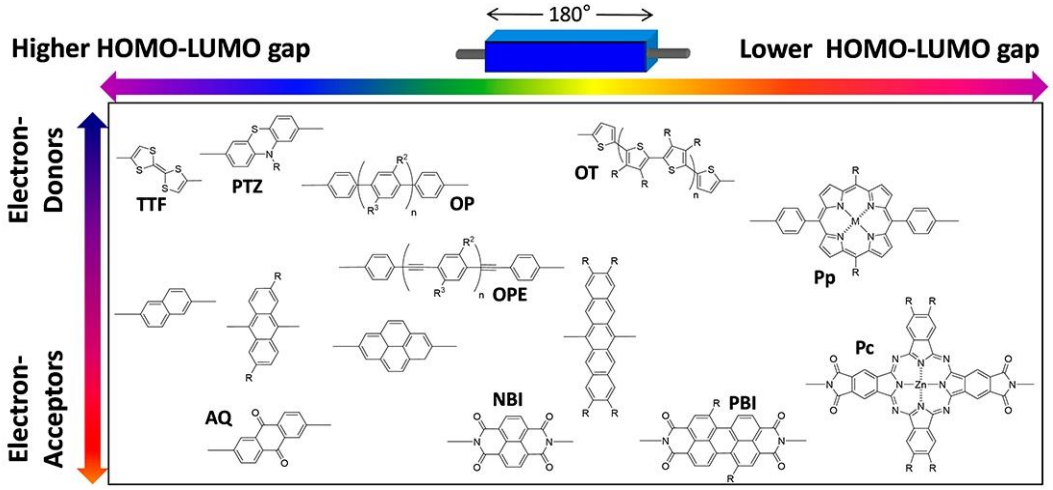


Figure 29. Potential π-conjugated functional blocks.

Ethynyl groups have been selected as the ***spacers*** between the different units forming the monomer. Such groups are interesting to us because they are linear, rigid and allow for substitution in this suitable 180º angle. They have some rotational liberty around the connecting σ-bonds and a minimum steric hindrance, which is necessary for sufficient conformational freedom between bases and central blocks. They are also π-conjugated and can facilitate electronic coupling between the different units in the cyclic tetramers. Finally, they possess a wide chemical versatility, mainly in simple and mild cross-coupling reactions,[120] and allow for a convenient coupling between central block and nucleobases through Sonogashira reactions, which are central in this project. The nucleobases could also be directly connected without intercalating any central block, through an ethynyl spacer or just a single covalent bond.

Overall, the linear structure of the dinucleobase monomer together with the 90º angle imposed by Watson-Crick H-bonding interactions at the edges will lead to the formation of cyclic, rectangular assemblies composed of four molecules.

3) Ongoing Projects in MSMn.

As stated before, this cyclic tetramer motif is central to several ongoing projects in the group and the contents of the current Thesis are just some of them. Since of all these projects are highly interrelated and have or are being developed in parallel, sometimes in collaboration

[120] R. Chinchilla, C. Nájera, *Chem. Rev.* **2014**, *114*, 1783–1826.

with other researchers in the group, we include herein a brief summary of some of them, which in most cases constitute the main body of just finished or ongoing Thesis.

1. *Study of the H-bonded Cyclotetramerization Process of Dinucleoside Molecules in Solution*. This is the first research line initiated in the group within this general scheme and the main results obtained so far are described in the current Thesis. Here, we wanted to study, in as much detail as possible, the process of cyclic tetramer self-assembly by Watson-Crick H-bonding between complementary bases in solution, using simple, properly designed dinucleoside monomers.

2. *Self-Assembled Photoactive Donor-Acceptor Systems*. The main purpose of this second research line is to introduce "function" in the central blocks between the nucleobase directors. Such function comes from the electronic properties of the π-conjugated semiconducting blocks installed in the central part of the monomer, as illustrated in Figure 29. This research line is clearly aiming at the development of nanostructured self-assembled materials for optoelectronic applications, in which molecular order at the nanoscale would result in improved device performance. However, the group soon realized that the incorporation of specific dyes into the monomer structure that can give rise to resonance energy transfer events between them, represented a useful way to study as well cyclotetramerization processes, self-sorting phenomena (see *Chapter 4*), and multicomponent macrocyclic assemblies (see *Chapter 6*).

3. *Nanotubular Systems Self-assembled through Orthogonal Supramolecular Interactions.* The main goal in this line is to investigate the possibility of using the cyclic tetramer as a monomer in supramolecular polymerization processes. Being a cyclic structure, we expect the polymer to be tubular, that is, a self-assembled nanotube. In order to drive and guide the polymerization process, the group has studied the role of *parallel directors*. These are simple amide or urea substituents, or peptide fragments that typically form beta-sheets in structural proteins, which can polymerize parallel to the stacking axis and that will be strategically placed at the periphery of the molecules. This polymerization process is being studied in both apolar organic solvents and in water, by respectively equipping the monomers with lipophilic or water-solubilizing chains. Furthermore, if amphiphilic central blocks are incorporated in the monomer, one may reach a control over the functionality of the central nanotube pore, in order to make it either hydrophilic or hydrophobic, so as to control guest uptake.

4. *Tubular Quadruplex DNA*. Taking the cyclic tetramer structure as a model, the group plans to radically alter DNA self-assembly at the deepest level: the interaction between bases. For such a goal, the pyrimidine moiety in dideoxynucleoside monomers, related to the ones described in Figure 27, will be properly functionalized to react them in a DNA synthesizer through phosphoramidite chemistry, so as to prepare length- and sequence-tailored oligonucleotides. Since each base is substituted with the complementary one in an adequate geometry, the spontaneous self-recognition of 4 oligonucleotide strands *via* Watson-Crick interactions is expected to yield a quadruplex DNA structure, having a cyclic section and a chiral, tuneable inner pore.

5. *Molecular Recognition onto Nanostructured Surfaces*. The group studied how to transfer all the rich and versatile supramolecular macrocyclization chemistry onto surfaces, so as to

characterize and study our cyclic tetramer assemblies by Scanning Tunneling Microscopy (STM) and other surface-based techniques. Three different deposition techniques and interfaces were employed to study the 2D self-assembly process: 1) *solid-liquid* interface, after drop-casting deposition onto Highly Ordered Pyrolytic Graphite (HOPG);[121] 2) *solid-vacuum* interface, by sublimating the molecules under ultra-high vacuum and depositing them on metallic substrates, such as Ag (111) or Au (111), by molecular beam epitaxy, and 3) *solid-air* interface *via* the preparation of Langmuir-Blodgett films. The final goal would be to obtain porous nanostructured surfaces that are able to recognize and host specific molecules as a function of their size, shape and chemical structure (Figure 30). Most of the work developed in this project constitutes the core of a finished Thesis (N. Bilbao, *"Porous 2D Nanostructured Networks via Nucleobase Self-Assembly"*, February 2016). Due to the similarity of the objectives and the relevance of the work, especially when employing the HOPG-solution interface, which provided higher versatility and control, we are summarizing in the next section the most important results of the mentioned Thesis that constitute an important background to the current one. More recent research in this line is now focused to the generation of H-bonded organic frameworks.

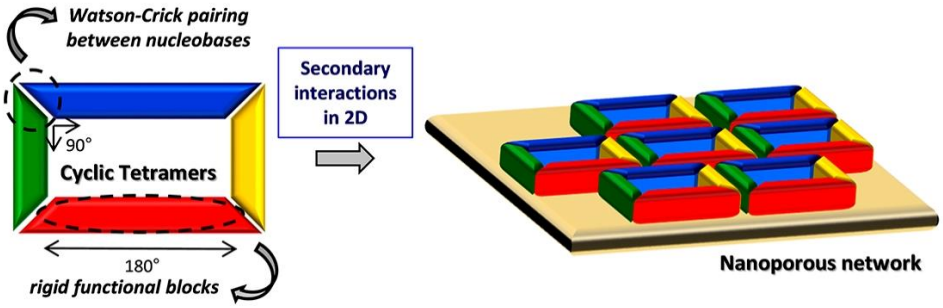


Figure 30. Self-assembly strategy toward nanostructured self-assembled porous networks.

4) Cyclic Tetramer Self-Assembly onto Surfaces.

- Cyclic tetramer network formation. Despite the designed monomers are preorganized for ring closure, the intramolecular event that shifts equilibria toward cyclization is not necessarily favoured when the molecules are concentrated on the surface. Different generations of both GC and AU monomers were designed and synthesized. In this way, stable networks of cyclic tetramers, H-bonded through their Watson-Crick units, were obtained after sample preparation optimization. This stabilization is caused by partial occupation of the inner pore of the tetramers by alkoxy chains, as well as by secondary interactions established between the adsorbed tetramers. In the GC structures, H-bonding between aminopyridine fragments occurred, so the tetramers were bound through their G edges to form the network (Figure 31a,b). In contrast, in the AU system, secondary H-bonding interactions were established between the external U carbonyl lone pair and A amine proton that were not participating in the Watson-Crick pairing, so the AU tetramers were bound through their corners (Figure 31c,d).

[121] N. Bilbao, I. Destoop, S. De Feyter, D. González-Rodríguez, *Angew. Chem. Int. Ed.* **2016**, *55*, 659–663.

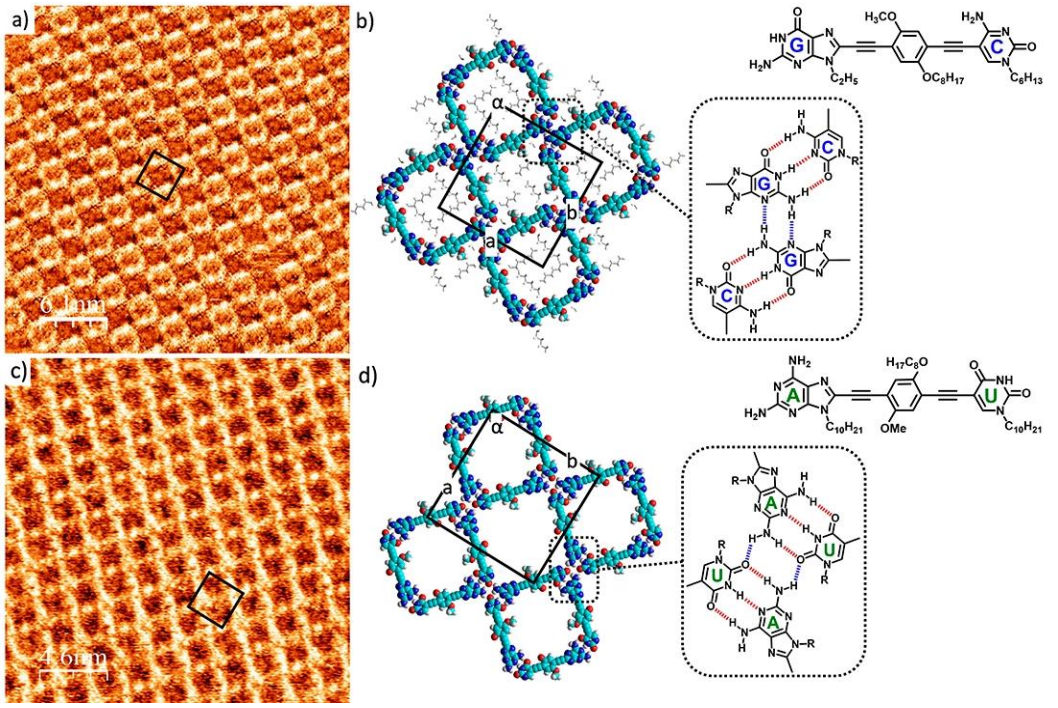


Figure 31. *Self-assembled porous networks formed by GC3 and AU2 on HOPG from solutions in TCB:OA (1:1).* (**a**) STM image of **GC** (8.2 x 10^{-6} M; I_{set} = 200 pA, V_{bias} = -300 mV). The unit cell is indicated by black lines (a = 3.6 ±0.1 nm, b = 3.6 ±0.1 nm, α = 89 ±1°). (**b**) Proposed model for **GC**, along with chemical structure of the compound and scheme of the stabilizing motif. (**c**) STM image of **AU** (7.9 x 10^{-5} M; I_{set} = 100 pA, V_{bias} = -350 mV). Unit cell parameters: a = 3.5 ±0.1 nm, b = 3.5 ±0.1 nm, α = 91 ±1°. (**d**) Proposed model for **AU**, along with chemical structure of the compound and scheme of the stabilizing motif. Dashed red lines indicate Watson-Crick base-pairing and dashed blue lines correspond to secondary H-bonding between nucleobases.

- Host-guest chemistry. A central cavity is generated when the tetramers are deposited onto surfaces. The possibility to fill these cavities with a suitable guest was therefore studied next. For such purpose, coronene was added and the ability of the nanoporous densely-packed network to host size-complementary guests was proved (Figure 32). Larger guests like phthalocyanines were however not able to fit within the cavities and their co-assembly on the surface was not observed.

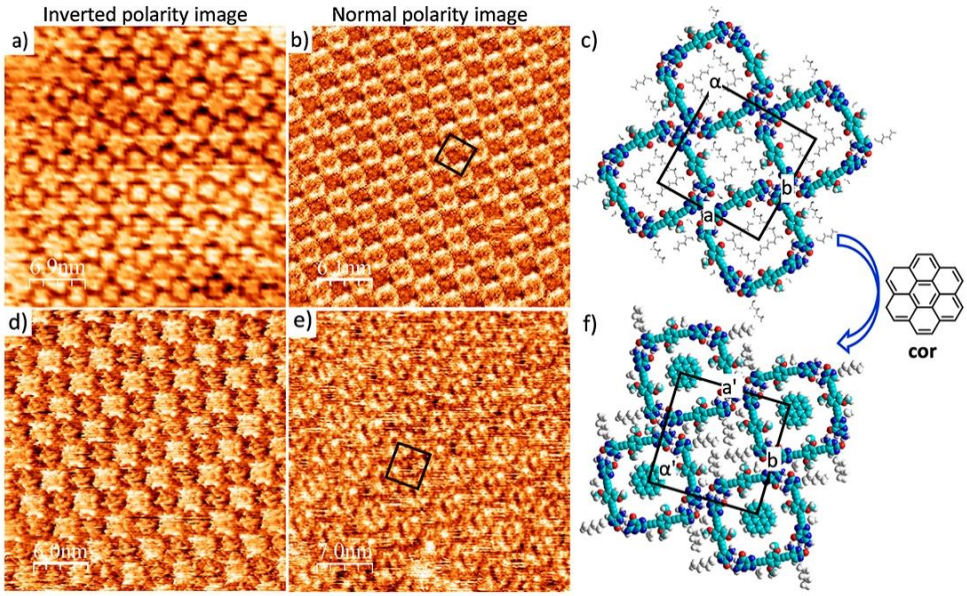


Figure 32. *Self-assembly of the GC3 one-component and GC3-cor systems at the HOPG/TCB:OA (1:1)* **interface.** Normal polarity tunneling parameters: I_{set} = 80 pA, V_{bias} = -350 mV. Inverted polarity parameters: I_{set} = 80 pA, V_{bias} = 350 mV. (**a**) Inverted polarity STM image of a single **GC** domain. (**b**) Normal polarity STM image of **GC**. The unit cell is indicated by black lines (a = 3.6 ±0.1 nm, b = 3.6 ±0.1 nm, α = 89 ±1°). (**c**) Model proposal for the one-component system. (**d**) Inverted polarity STM image of **GC** and **cor** (1:200; 6.0 x 10^{-6} M). (**e**) Small scale normal polarity STM image of the **GC**-**cor** system (a' = 3.6 ±0.1 nm, b' = 3.5 ±0.2 nm, α' = 89 ±1°). (**f**) Model proposed for the host-guest system.

- Size control. Another interesting approach is the possibility to tune the length of the central block in order to change pore size and thus study the ability of the 2D network to host different guest molecules as a function of their size. As it is depicted in Figure 33, a shorter G-C monomer (Figure 33a) can self-assemble in cyclic tetramers at the solid-liquid interface too. A longer monomer bearing a bithiophene central block was also synthesized with the aim of obtaining larger cavities.

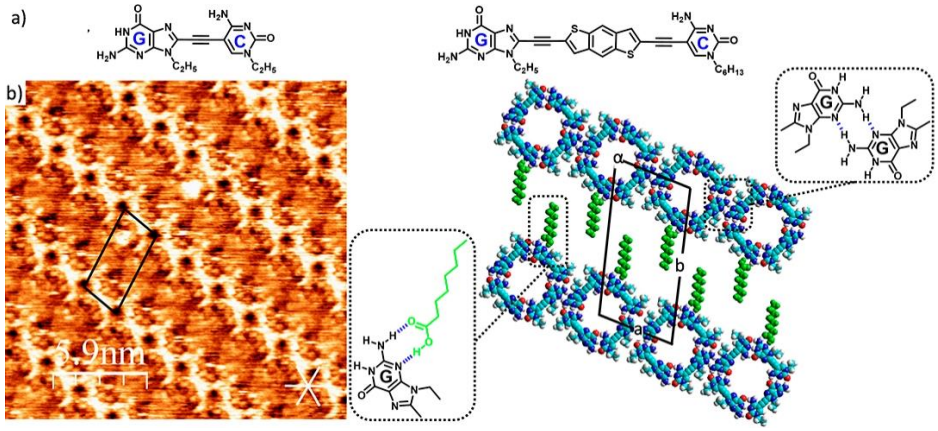


Figure 33. Pore size control and size-selective guest adsorption strategy. (**a**) Chemical structures of the short **GC** and the long **GC** monomers. (**b**) STM image of **GC** self-assembly at the HOPG/TCB:OA (1:1) interface (1.1 x 10^{-6} M; I_{set} = 60 pA, V_{bias} = -300 mV), along with the proposed association mode with schematics of the stabilizing H-bonding motifs. The unit cell is indicated by black lines (a = 2.5 ±0.1 nm, b = 4.6 ±0.1 nm, α = 106 ±1°). White lines indicate the normal axes of graphite.

- Self-sorting. In order to prove self-sorting phenomena in two dimensions, due to the specificity and complementarity of the H-bond DNA-nucleobases, a new series of GC and AU monomers of different length or π-electron density were co-deposited on the substrate in order to try to ascertain whether the domains of cyclic tetramers did not mix and remained formed by a single molecular component (Figure 34).[122] The preliminary STM images revealed a complex mixture of domains but also single domains corresponding to both small and large tetramers separately (Figure 34). Although the experimental conditions have not been optimized yet, these initial results were encouraging and this approach can be regarded as a valuable route to the future engineering of multicomponent nanostructures on a surface.

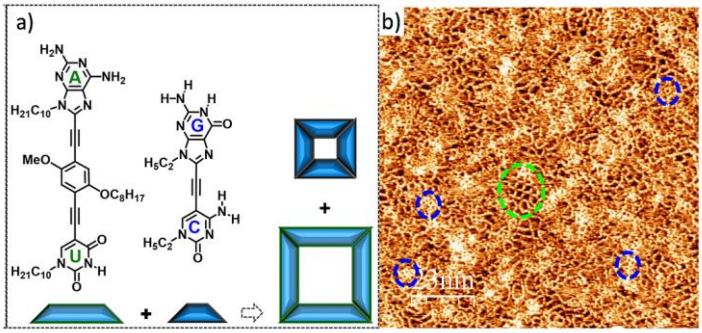


Figure 34. *Proposed self-sorting experiments* based on (**a**) size differentiation of dinucleobase monomers for the elucidation of self-sorting phenomena on the surface. (**b**) STM image of a 1:3 ratio (**AU**: 4.0 x 10^{-5} M and **GC**: 1.2 x 10^{-4} M) premixed solution in TCB/OA (1:1) on HOPG (I_{set} = 80 pA, V_{bias} = -300 mV) corresponding to self-sorting experiment (**a**). Blue dashed circles indicate small GC tetramer domains and green AU bigger macrocycle domain.

- Multicomponent and size control. This is the last topic addressed at the solid-liquid interface that is related to the contents of this Thesis. Symmetrically substituted monomers of different lengths were prepared for the construction of two-component cyclic species (ABAB; Figure 35) and the study of their 2D self-assembly. Upon H-bonding association of the respective monomers to a rectangular-shaped macrocyclic network, a second goal would be to investigate their shape-discriminating host-guest ability. A possible shape-matching guest would be the pentacene molecule, which can fit in the inner tetramer cavity, whereas a larger disk-shaped molecule such as coronene should not be included in this host network. This is a still unfinished work where a wide set of new molecules were carefully designed, synthesized and are ready to be investigated in the near future.

[122] Nerea Bilbao Bustinza, Doctoral Thesis, *"Porous 2D Nanostructured Networks via Nucleobase Self-Assembly"*, Universidad Autónoma de Madrid, February **2016**.

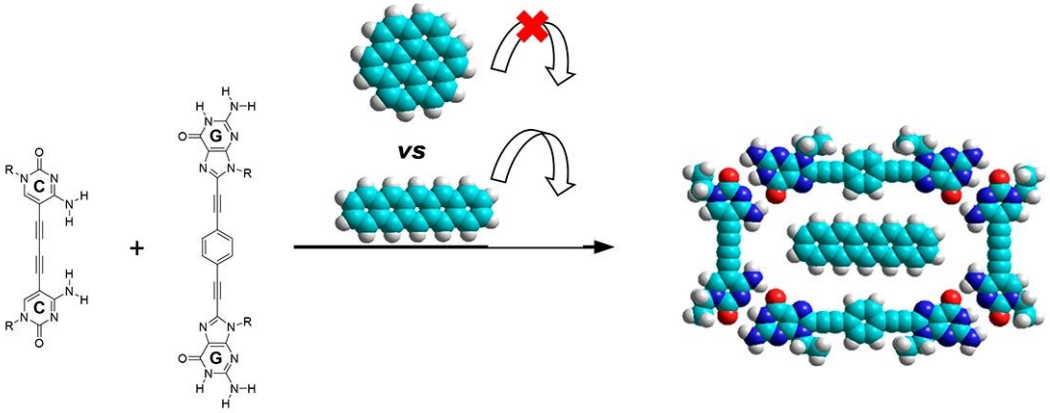


Figure 35. Proposed shape-discriminating experiment between ABAB tetramers, and pentacene as shape complementary guest.

Objectives.

This Doctoral Thesis aims to contribute to the field of the supramolecular chemistry and self-assembly with the study of *mono- and multicomponent cyclotetramerization processes in solution driven by H-bonding between complementary DNA bases*. Following the design guidelines explained in the previous section, our final, self-assembled products would be thus finely-tuned discrete tetrameric macrocycles. The supramolecular properties of nucleobases, in terms the selectivity and directionality of the different H-bonding patterns are therefore an essential tool in our scheme. In short, in this Thesis we aim to ***develop an unconventional and versatile strategy based on complementary H-bonding toward discrete self-assembled macrocycles,*** *as well as to* ***study the underlying supramolecular properties involved in the cyclotetramerization processes.***

The formation of these discrete H-bonded nanostructures will be analyzed by different 1D and 2D ^{1}H NMR techniques and methods (titrations, concentration- and temperature-dependent measurements, Diffusion-Ordered Spectroscopy (DOSY), etc.), mass spectroscopy (ESI Q-TOF), X-ray diffraction and, in a complementary way, absorption, emission and CD spectroscopies.

This thesis presents different objectives, distributed in *Chapters 1-6*, in which we analyze different parameters and gradually increase the degree of cyclic tetramer complexity:

The first objective of this Thesis (*Chapter 1*) is the design and synthesis of the respective molecular components that will constitute the target monomers. Firstly, we aim to develop optimized synthetic routes to a wide family of dihalogenated central blocks that will comprise the core of the monomer, as well as to a complete variety of nucleobase derivatives substituted at one end with an ethynyl moiety. We will explore the reactivity of nucleobase derivatives appropriately functionalized at the **R^{1}** position (see Figure 36) and optimize the synthetic routes leading to them, some of them previously reported by other researchers, in terms of overall yield, convergence, and synthetic convenience. Finally, a convergent coupling route, based on palladium-catalyzed Sonogashira reactions, toward the target monomers, in which the chosen central block and H-bonding directors are incorporated, will be pursued.

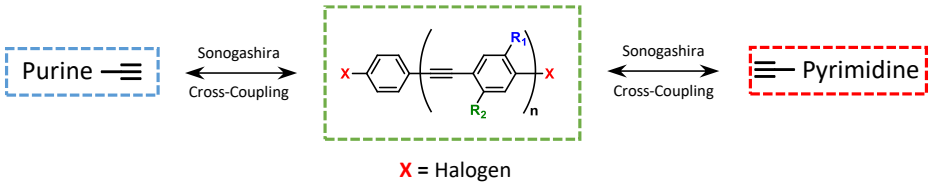


Figure 36. Molecular components considered and synthesized in *Chapter 1.*

The second objective (*Chapter 2*) is the determination of the reference association and dimerization constants of the different H-bonded nucleobase-pairs in diverse solvent systems. These values, many of which had not been determined before in the literature, will be subsequently applied in the study of the macrocyclization processes. For such a goal, we will synthesize a new family of lipophilic nucleobases linked to a short π-conjugated oligophenylene-ethynylene moiety, resembling as much as possible the π-conjugated structure of our target monomers (Figure 37). Also, the **R^{1}** position will be appropriately functionalized with a ribose moiety with bulky substituents in order to prevent π-π staking

and afford characteristic signals in ^{1}H NMR, which will be common in all the monomers synthesized in this Thesis.

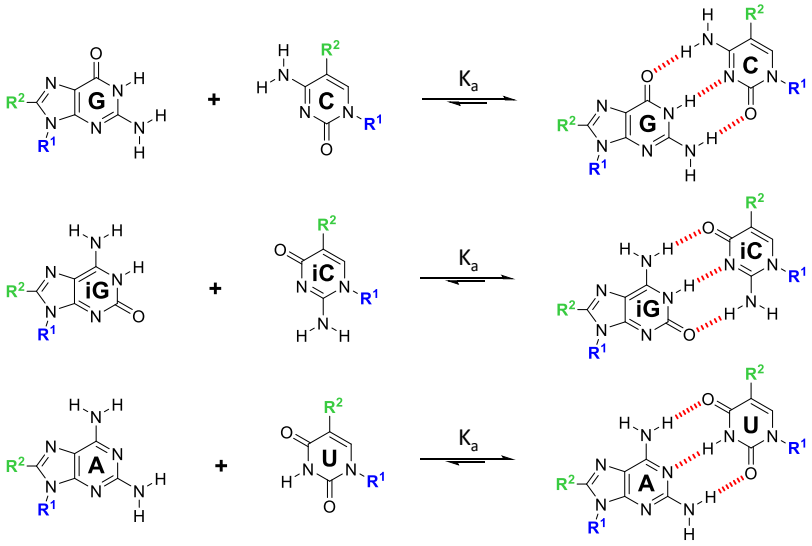


Figure 37. Equilibria between complementary nucleobases.

In *Chapter 3*, we will examine the fidelity of the cyclotetramerization process or, in other words the association selectivity in discrete cyclic tetramers, instead of open oligomeric structures or other kind of strained cycles. For this purpose, a series of three ditopic molecules, carrying complementary nucleobase derivatives on each side (G-C, iG-iC and A-U), will be prepared and studied. These monomers have been designed to optimize the *Chelate Effect*, and therefore high *EM* values are expected, which will be calculated by means of different methods and techniques. Furthermore, we will try to determine the role of the nature of the multipoint H-bonding interaction patterns on the chelate cooperativity of these supramolecular macrocyclization processes (Figure 38).

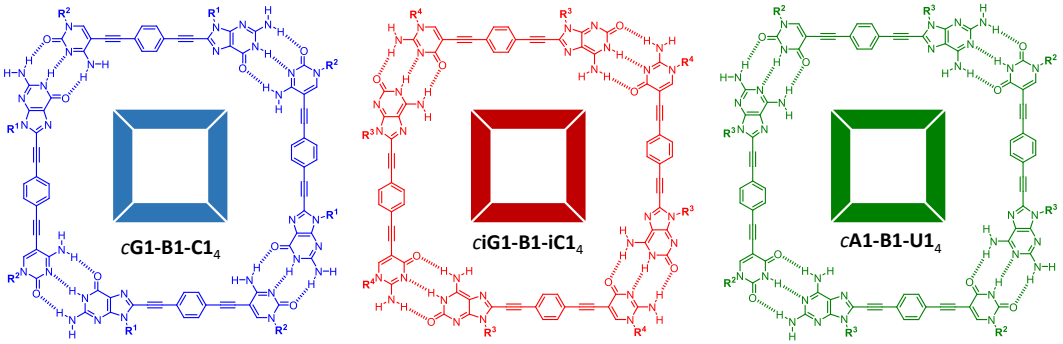


Figure 38. One-component cyclic tetramers formed by the H-bonding association of complementary dinucleoside monomers.

Addressing self-sorting phenomena is the main topic of *Chapter 4*. In particular, the narcissistic macrocyclization behaviour of our three ditopic complementary monomers will be analysed (Figure 39). Because of the different symmetries (*ADD-DAA* vs *DAD-ADA*) of the H-bonding patterns of the G-C + A-U and iG-iC + A-U combinations, we expect these dinucleoside

monomers to self-sort, that is, to self-associate independently in their respective cyclic tetramers. However, in the case of the G-C + iG-iC combination, the nucleobase pairs share similar unsymmetric H-bonding patterns, and both regular (G: C or iG:iC) and reverse (G:iC or iG:C) Watson-Crick H-bond pairs can be formed with absent selectivity. Hence, the combination of these two molecules should lead to a complex mixture of open and cyclic structures and self-sorting should not operate based solely on nucleobase H-bonding selectivity. Only if other factors come into play, like chelate cooperativity effects, self-sorting would dominate.

Figure 39. Expected self-sorting phenomena of the ditopic monomers.

Once the cyclotetramerization process will be thoroughly studied with a model monomer structure, in *Chapter 5* we will tune the size of the macrocycle by simple variation of the length of the central block, while maintaining the same G-C nucleobase pairing. As it is shown in Figure 40, this can be controlled by regulating the length of the central blocks placed between G and C nucleobases, from a simple C-C bond to a series of phenylene-ethynylene oligomers. The nature and flexibility of the central block can have important consequences on the magnitude of *EM*, so our aim in this Chapter is to estimate the length limit of the central block that is needed to obtain quantitative cyclic tetramer assembly in different solvents.

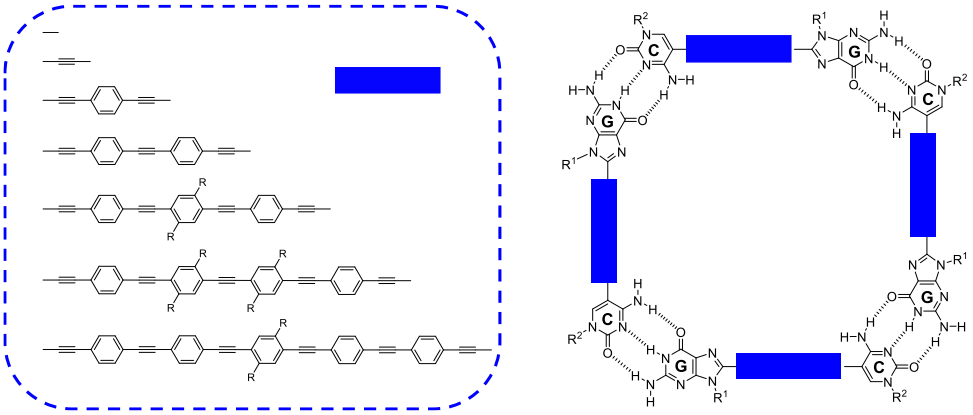


Figure 40. Central blocks with different lengths between complementary G-C base-pair.

Ongoing Objectives

As a final objective of this Thesis, we targeted to reach a new additional level of complexity forming in solution multicomponent tetramers (ABAB). These multicomponent systems will be prepared using binary combinations of the appropriate monomers disubstituted by non-complementary nucleobases (Figure 41).

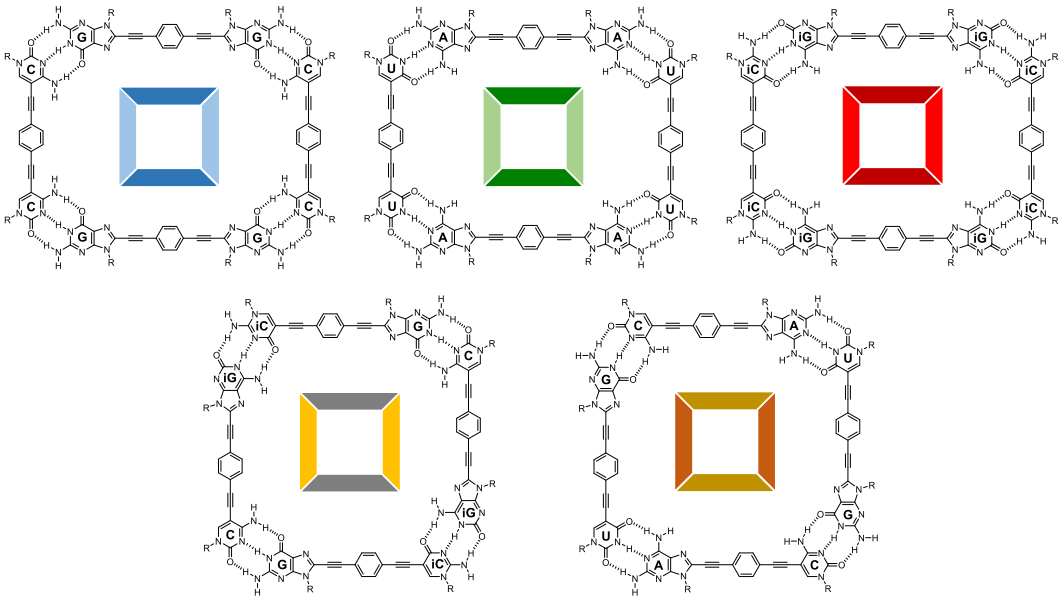


Figure 41. Different ABAB systems proposed in this Thesis.

Finally, we would like to comment briefly on two other objectives that were initially targeted at the beginning of this Thesis, but that could not be concluded due to the limited time and to their inherent complexity, so only a few details are included in this manuscript. These objectives denote the extraordinary robustness and versatility of the macrocyclization strategy followed in this Thesis and clearly constitute highly ambitious projects in the field of *Supramolecular Chemistry*.

One of them focuses on continuing the expansion of cyclic tetramer complexity to systems formed by three (ABAC) and four (ABCD) different units (Figure 42). A novel family of ditopic monomers, having carefully designed nucleobase combinations at their edges so that a single cyclic tetramer can be formed, need to be prepared for this purpose. We expect chelate cooperativity to be reduced significantly as the number of components increases, and therefore the ribose moiety will be substituted with long alkyl chains in order to afford solubility in non-polar solvents and increase Watson-Crick binding strength. Two possible monomer combinations are shown in Figure 42.

- The *3-component system* (ABAC; Figure 42a) is also interesting to the group because it gives the possibility of positioning different functional units (for instance, energy donors and acceptors, p- and n-type semiconducting molecules, etc.) at defined distances.
- The *4-component* (ABCD; Figure 42b) tetramer is the most complex system attainable and its formation requires a third nucleobase pair, such as the iG-iC couple.

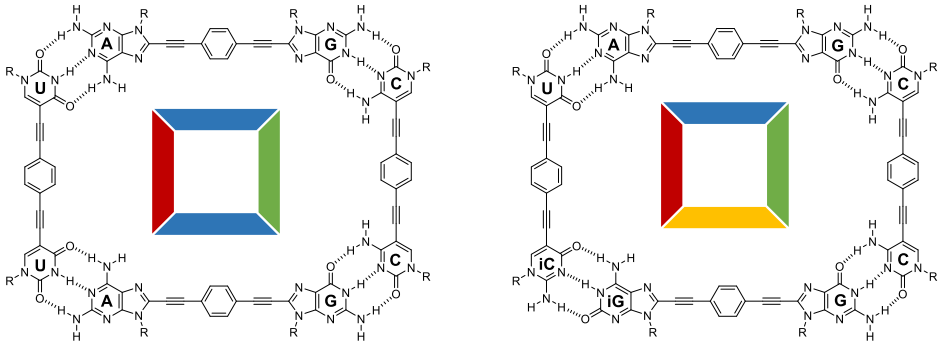


Figure 42. ABAC systems proposed in this Thesis.

The last topic is devoted to leave the 2D monocyclic level and reach the 3D multicyclic level. Concretely, we are aiming to the production of discrete prisms of defined geometries, which can be nowadays easily prepared *via* metal-ligand coordination approaches (see Introduction), but have, to the best of our knowledge, never been targeted through H-bonding interactions. For such a goal, we will prepare a new family of π-conjugated central blocks with different geometric characteristics depending of the desired prism (Figure 43):

- Linear *p*-disubstituted oligomer combinations would form *rectangular cyclic tetramers* (as described above).
- The same linear oligomers, combined with 1,3,5-trisubstituted arenes, *meso*-tetrasubstituted porphyrins, or hexasubstituted arenes would give rise to the formation of *trigonal, tetragonal or hexagonal prisms*, respectively.
- *meso*-Tetrasubstituted porphyrin combinations may, on the other hand, generate *cubes*.

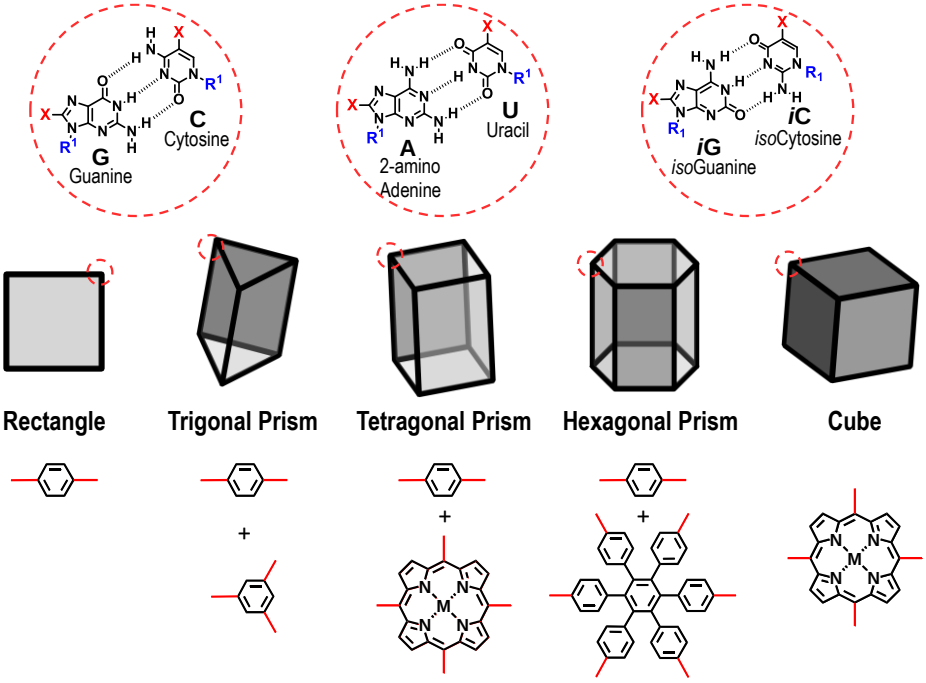


Figure 43. Structure of the different prisms synthesized in this Thesis with their respective central blocks.

Unfortunately, this work is still in progress and although most molecular units have been already synthesized, not all the experiments by NMR techniques have been carried out yet. However, the preliminary results suggest the interaction of the respective complementary nucleobases.

Chapter 1

Monomer Design and Synthesis

This Chapter is devoted to synthesize the targeted molecules which form part of the final monomers. These molecules comprise different derivatives of pyrimidines (cytidine, uridine and *iso*cytosine), purines (guanosine, 2-aminoadenosine and *iso*guanosine) which form the self-recognition moieties, and finally, a new series of oligophenylene-ethynylene that will tune the distance between nucleobases and therefore, the tetramer size.

1.1. Monomer Design and Synthetic Strategy.

The monomer design and key structural elements to form cyclic tetramer species in solution has been explained before in the Background Section. This Thesis will keep the basic monomer design but the nucleobase substituents and the central blocks will suffer some modifications. The monomers will be constituted by rigid and planar π-conjugated central blocks substituted at both edges with nucleobase directors with an exact angle of 180°. Watson-Crick H-bonding interactions between complementary nucleobases will then provide the required 90° angle to yield rectangular closed assemblies (Figure1.1). In this first chapter, we will independently describe the molecular design and synthetic preparation of the different molecular components: nucleobase directors and central blocks.

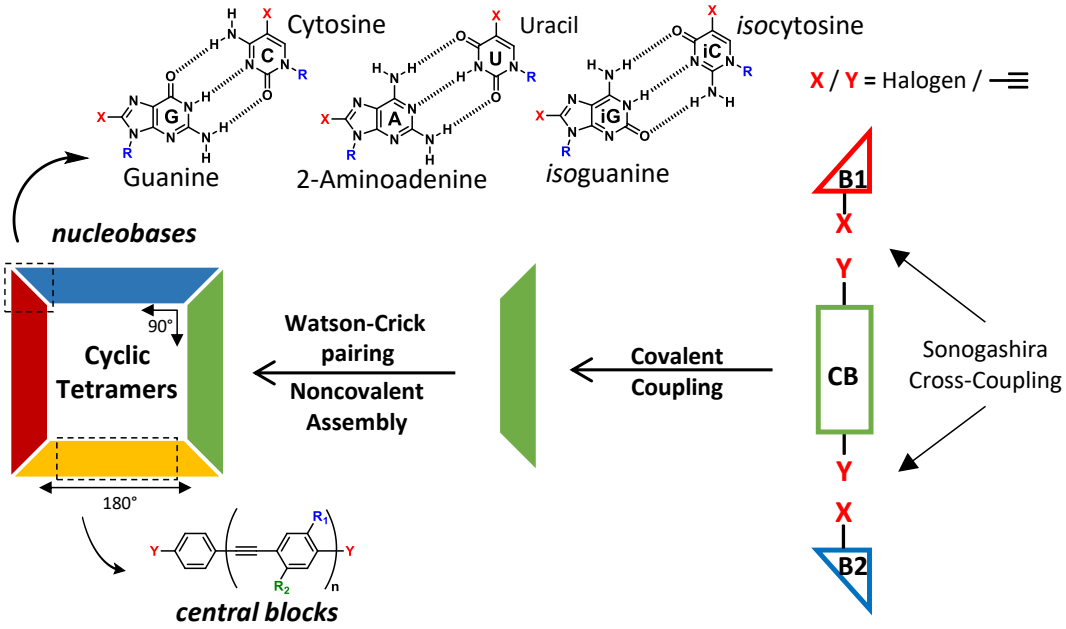


Figure 1.1. Molecular components and self-assembly strategy toward cyclic tetramers. Central Block (CB), Nucleobase (B).

Central block derivatives have been properly chosen taking into account different issues: (1) the requirement of linear substitution at both edges; (2) the ability to tether specific lateral groups that can supply solubility to the assemblies; and (3) the possibility of tuning their length, which will at the same time regulate cyclic tetramer size. This last issue will be specifically addressed in *Chapter 5*. Therefore, a series of oligophenylene-ethynylene central blocks, having from 1 to 5 phenyl groups, will be used to modulate tetramer size.[123] Due to

[123] To the sake of simplicity, we have labelled each central block according to the number of aromatics rings that form the molecule.

the increment in π-conjugation, the monomer can absorb light in the visible region, which will allow us to follow the macrocyclization processes by spectroscopic and ^{1}H NMR techniques. The use of specific dyes as central blocks can afford additional interesting possibilities, especially if energy donor-acceptor pairs are employed. Although this is not the objective of this Thesis, parallel work in our laboratories focused on the use of FRET pairs as central blocks for the study of different supramolecular processes, such as, as will be discussed in *Chapter 4*, self-sorting phenomena.

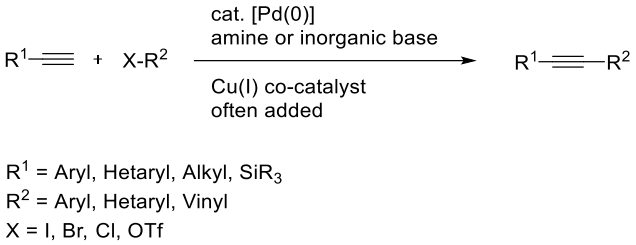
The ribose unit at the ***nucleobases*** will be equipped with bulky lipophilic groups. Since the main objective in this Thesis is the study of the cyclotetramerization process from dinucleoside precursors, these groups are useful to avoid π-π stacking between planar tetramers, afford an adequate solubility in non-polar organic solvents, and provide characteristic signals in ^{1}H NMR that facilitate the analysis and identification of the different supramolecular species.

Regarding the ***spacer*** moiety connecting nucleobases and central blocks, the choice in this overall self-assembly scheme is the ethynyl group. This group, comprising a sp-sp carbon-carbon triple bond, displays four different characteristics that are important for our purposes: it is linear, rigid, π-conjugated, and presents low steric volume. These attributes are essential to connect properly central blocks and nucleobase derivatives without causing important steric effects. The ethynyl group is also synthetically convenient. It can be easily incorporated through Sonogashira cross-coupling reactions between two different, appropriately substituted fragments.

1.1.1. Cross-Coupling between Nucleobases and Central Blocks: The Sonogashira Reaction.

In this Thesis, Sonogashira coupling reactions[124] have a special key role to synthesize our target monomers from its precursors: nucleobase derivatives and central blocks. For the formation of sp^2 / sp carbon-carbon bonds, coupling reactions catalyzed by palladium are one of the most commonly used methodologies. In particular, Sonogashira cross-coupling reaction are used to synthesize natural products, heterocycles, molecular electronic units, conjugated polymers or nanostructures; and many examples are available in the literature.[125] The Sonogashira reaction consists in a Pd-catalyzed C-C bond formation process which can bond a sp^2 carbon of an aryl or vinyl halide (or triflate) with a terminal sp hybridized carbon (Scheme 1.1).[126]. This process which can be easily carried out at room temperature with a palladium catalytic source combined with a co-catalytic amount of CuI in an amine reductive solvent, was proved first in 1975 by the Japanese Sonogashira and his colleagues Tohda and Hagihara

[124] a) R. Chinchilla, C. Nájera, *Chem. Soc. Rev.* **2011**, *40*, 5084–5121; b) R. Chinchilla, C. Nájera, *Chem. Rev.* **2007**, *107*, 874–922; c) H. Doucet, J.-C. Hierso, *Angew. Chem. Int. Ed.* **2007**, *46*, 834–871.
[125] a) L. A. Agrofoglio, I. Gillaizeau, Y. Saito, *Chem. Rev.* **2003**, *103*, 1875–1916; b) R. Bielski, Z. J. Witczak, *Chem. Rev.* **2013**, *113*, 2205–2243.
[126] K. Sonogashira, Y. Tohda, N. Hagihara, *Tetrahedron Lett.* **1975**, *16*, 4467–4470.



Scheme 1.1. Scheme for a general Sonogashira cross-coupling reaction.

Nowadays, it is assumed that two independent catalytic cycles form part of the reaction (Figure 1.2).[3b] The catalytic pathway starts with the "palladium-cycle" and the active species Pd^0L_2, which can be generated from Pd^0 complexes or Pd^{II} from complexes such as $Pd(PPh_3)Cl_2$, via formation of a $[Pd^{II}L_2(C{\equiv}CR^2)_2]$ species. The latter gives $[Pd^0L_2]$ after reductive elimination by forming a diyne. Amines may also reduce Pd^{II} to Pd^0 through formation of iminium cations.[127] Once complex $[Pd^0L_2]$ has been formed, the first step in the catalytic cycle is initiated by oxidative addition of the aryl or vinyl halide, which is considered to be the rate-limiting step of the Sonogashira reaction. The barriers of oxidative addition of ArX (X = Cl, Br, I) increase in the order of ArI < ArBr < ArCl, as it is believed to be an electron-donating step.[128] The formed $[Pd^{II}R^1L_2X]$ adduct is then transformed into a $[Pd^{II}L_2R^1(C{\equiv}CR^2)]$ species after transmetalation with a copper acetylide formed in the "copper-cycle" (cycle B). This adduct suffers reductive elimination to the final alkyne, regenerating $[Pd^0L_2]$.

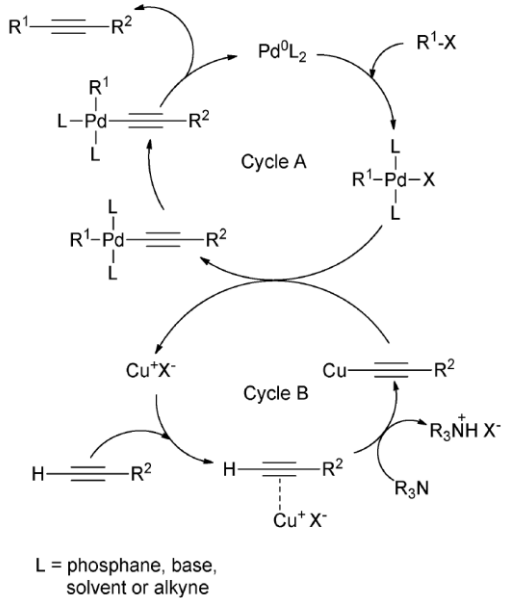


Figure 1.2. Supposed mechanism for the copper-cocatalysed Sonogashira reaction.[3a]

Due to the copper-mediated transmetalation of alkynes used in the catalytic cycle, is necessary to deoxygenate before the solvent of the reaction through freeze-pump-thaw cycles, in order to avoid unwanted alkyne homocoupling products. If oxygen is still remaining

[127] H. Li, G. A. Grasa, T. J. Colacot, *Org. Lett.* **2010**, *12*, 3332–3335.
[128] C. Gottardo, T. M. Kraft, M. S. Hossain, P. V. Zawada, H. M. Muchall, *Can. J. Chem.* **2008**, *86*, 410–415.

in the reaction media, Pd^0 complexes can be oxidized and placed out of the catalytic cycle. The alkyne derivatives could then be consumed forming homocoupling byproducts during the Sonogashira reaction. Their formation is explained by the Hay/Glaser reaction.[129]

In order to reach the target monomers through the most straightforward and high-yielding route from nucleobases and central blocks, two different Sonogashira pathways were evaluated (Figure 1.3):

(a) The ethynyl groups are placed in the central block and the halogen atom in the nucleobase precursor.

(b) The ethynyl group is placed in the nucleobase, and the halogen atoms at opposite sides of the central blocks.

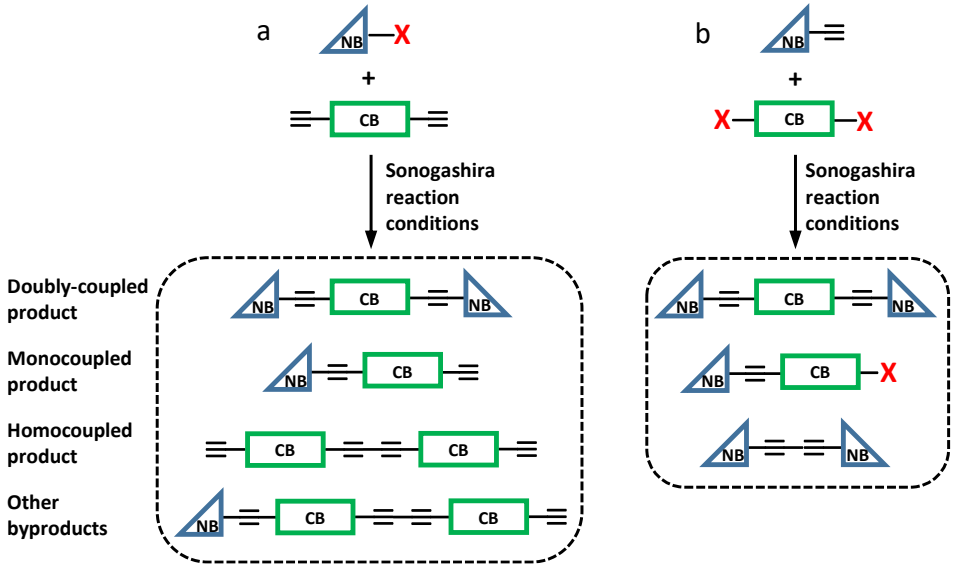


Figure 1.3.Two possible pathways to obtain the target monomers from central blocks (CB) and nucleobases (NB) *via* Sonogashira cross-coupling. X means halogen atom.

Three different groups of products are typically obtained in these reactions: 1) the doubly-coupled product that can be used to from ABAB systems (see *Chapter 6*); 2) the monocoupled product that can be used to form unsymmetrically substituted monomers in a subsequent coupling step, and 3) undesired homocoupled products, which are formed in minor amounts.

Although both routes **(a)** and **(b)** are effective and in some cases one must be chosen over the other for practical reasons, the experience accumulated in the group in general, and in this Thesis in particular, made us choose route **(b)** as the most convenient for our purposes. The route was more convergent and all the products could be purificated more easily. The reason is that only one homocoupled product, comprising two nucleobases, can be formed through route **(b)**, which is relatively easily differentiated and separated from the other reaction products. Besides, like the doubly-coupled product, the dinucleoside homocoupled product obtained through route **(b)**, having a diacetylene union, can be useful for the preparation of ABAB systems, so their isolation is typically interesting. In contrast, in route **(a)**

[129] G. Evano, N. Blanchard, M. Toumi, *Chem. Rev.* **2008**, *108*, 3054–3131.

different useless homocoupled byproducts can be formed in minor amounts and some of them are tricky to separate from the target product(s) due to their similar polarity.

Due to our preference for route **(b)**, we will describe hereafter the synthesis of dihalogenated central blocks, on one hand, and 5-ethynyl-pyrimidine or 8-ethynyl-purine nucleobase derivatives, on the other.

1.2. Synthesis of the Monomer Components.

The synthetic pathway of the different halogenated functional central blocks and the ethynylated nucleobase derivatives will be explained in this section. The final coupling between all the components obtained previously through consecutive Sonogashira reactions to generate the target monomers will be described separately in the corresponding Chapters, because of the existence of different strategies to synthesize unsymmetrically and symmetrically substituted derivatives.

1.2.1. Central Blocks.

A series of seven dihalogenated aromatic molecules have been chosen as central blocks (Figure 1.4). All of them are based in oligophenylene-ethynylene moieties of different lengths in order to regulate the monomer size while keeping the necessary rigidity in this part of the molecule. Furthermore, these derivatives can be equipped with large alkoxy chains (chiral or not) that will help to solubilize the monomers having longer aromatic bodies, which tipically have a stronger tendency to aggregate by π-π stacking.

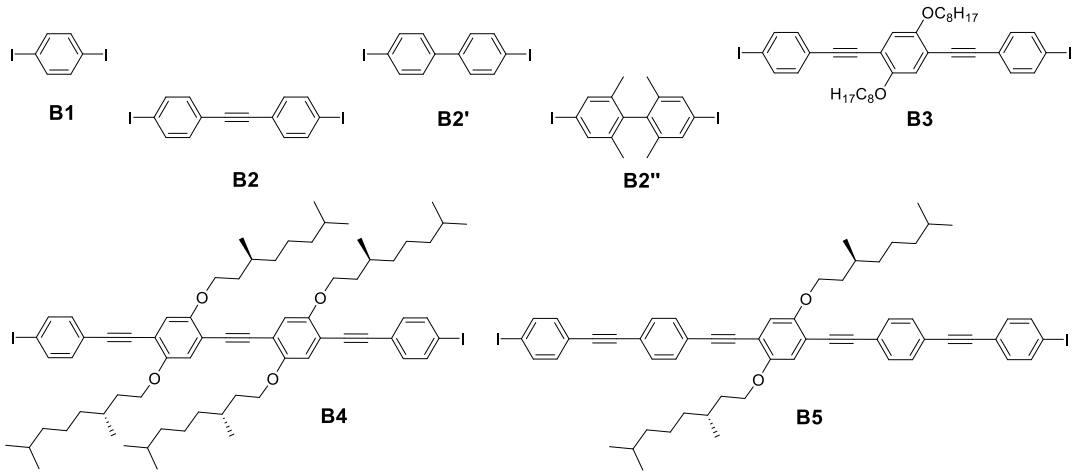
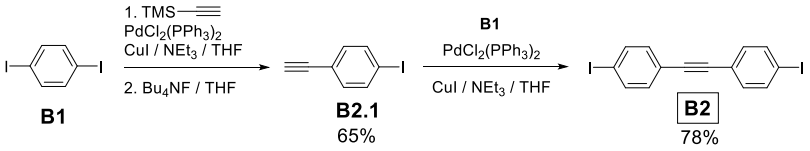


Figure 1.4. Targeted dihalogenated blocks **B1**-**B5.**

1,4-Diiodobenzene (**B1**) and 4,4'-diiodobiphenyl (**B2'**) were purchased from commercial suppliers.

B2. 1,2-bis(4-iodophenyl)ethyne (**B2**) was synthesized through a three-step synthetic route from commercial 1,4-diiodobenzene (**B1**) as it is shown in Scheme1.2.

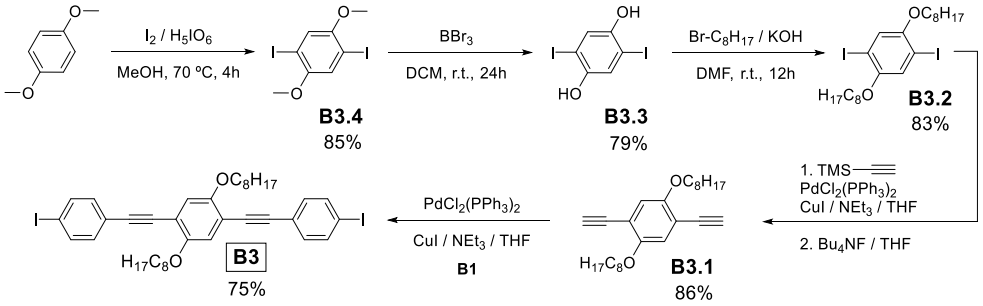


Scheme 1.2. Synthetic route to 1,2-bis(4-iodophenyl)ethyne (**B2**).

Sonogashira cross coupling with trimethylsilylacetylene (TMSA) and subsequent deprotection with Bu_4NF (TBAF) gives 1-ethynyl-4-iodobenzene (**B2.1**). Then, a new Sonogashira cross-coupling was set up with **B1** in order to obtain the desired product **B2** in good yields. These reaction conditions were optimized by using a large excess of the starting

material **B1** in both reactions, in order to favour the monocoupling derivative instead of the symmetrically substituted one. The excess of **B1** was recovered. The similar polarity of all the products forced us to purify them by means of long chromatography columns eluted with cyclohexane.

B3. Central block **B3**[130] was equipped with two symmetrical octyloxy chains to favour the solubility in organic media and to prevent π-π stacking. Molecule **B3** was obtained through a six-step synthetic route from commercial 1,4-dimethoxybenzene as indicated in Scheme 1.3.

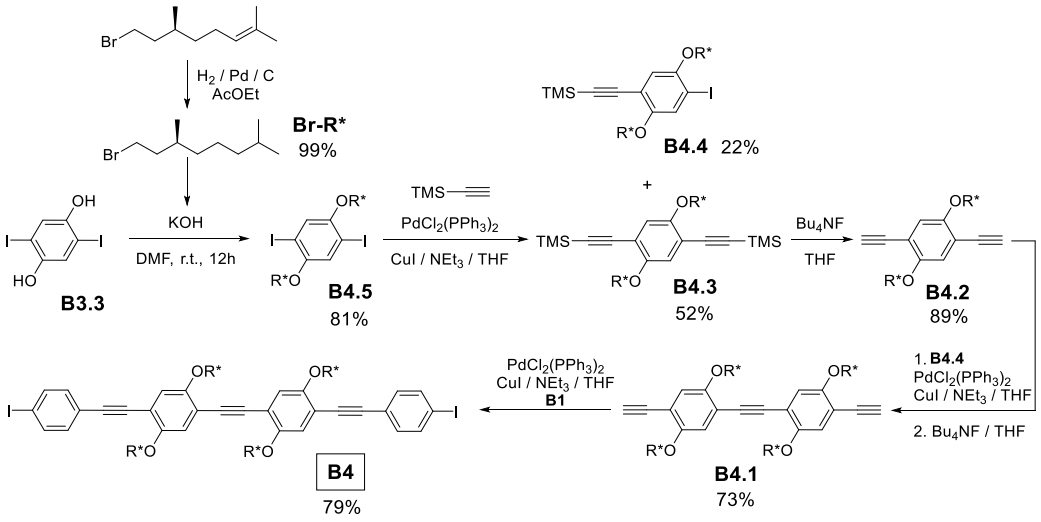


Scheme 1.3. Synthetic route to compound **B3**.

Selective iodination of 1,4-dimethoxybenzene yielded **B3.4** which, after deprotection of the methyl groups with BBr_3, afforded compound **B3.3**.[131] Then, a Williamson reaction in the presence of an excess of the alkyl chain gives the symmetrically alkylated product **B3.2**. A Sonogashira cross coupling in THF was then carried out with an excess of TMSA, followed by deprotection of the TMS group with TBAF to obtain **B3.1**. In order to synthesize the final product (**B3**), a large excess of **B1** over **B3.1** must be added: ten equivalents at least. The reason is the presence of two reactive positions for each molecule in Sonogashira cross-coupling conditions. With this large excess, the formation of **B3** is highly optimized and the generation of a complex mixture of undesired coupled derivatives avoided. One more time, the excess of **B1** was recovered.

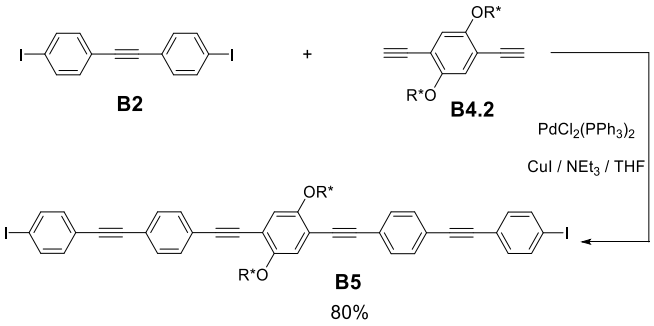
B4. Molecule **B4** was synthesized starting from the previously obtained **B3.3** derivative (Scheme 1.4). Firstly, the saturated chiral alkyl chain was obtained following a previously reported synthetic route.[132] Substitution at the 1,4 positions with the saturated chiral bromoalkane precursor, and then Sonogashira cross coupling with 1.5 eqs of TMSA gives **B4.4** and **B4.3**. This step was considered to avoid the generation of undesired products in the subsequent Sonogashira cross coupling reaction. Once **B4.3** was isolated, deprotection with TBAF afforded **B4.2**. A Sonogashira cross coupling reaction was carried out with an excess of **B4.4** in order to obtain the monocoupled derivative in high yield. Once again, deprotection with TBAF affords **B4.1** as a brown oil and a last Sonogashira cross coupling with a large excess of **B1**, gives **B4** with a good yield, as it is shown in Scheme 1.4.

[130] M. Schroeter, M. Behl, C. Weder, A. Lendlein, Mater. Res. Soc. Symp. Proc. **2012**, *1403.*
[131] a) S.-B. Ko, A.-N. Cho, M.-J. Kim, C.-R. Lee, N.-G. Park, *Dyes Pigm.* **2012**, *94*, 88–98; b) T. Shiraki, S. Haraguchi, Y. Tsuchiya, S. Shinkai, *Chem. Asian J.* **2009**, *4*, 1434–1441.
[132] M. M. L. Nieuwenhuizen, T. F. A. de Greef, R. L. J. van der Bruggen, J. M. J. Paulusse, W. P. J. Appel, M. M. J. Smulders, R. P. Sijbesma, E. W. Meijer, *Chem. Eur. J.* **2010**, *16*, 1601–1612.



Scheme 1.4. Synthetic route to compound **B4**.

B5. Compound **B5** was obtained by a single Sonogashira cross coupling reaction between **B4.2** and a large excess of **B2**. In these conditions, only the doubly coupled desired product was obtained in good yields, as shown in Scheme 1.5.



Scheme 1.5. Synthetic route to compound **B5**.

1.2.2. Nucleobase directors.

A series of chemically-modified lipophilic natural and non-natural nucleobase derivatives (Figure 1.5) that are equipped either with a halogen bond or an ethynyl group and substituted, in most cases, with a ribose moiety with different substituents have been synthesized.[133] These include cytidine (**C**), *iso*cytosine (**iC**) and uridine (**U**) as pyrimidine derivatives and guanosine (**G**), *iso*guanosine (**iG**) and 2-aminoadenosine (**A**) as a complementary purine bases.

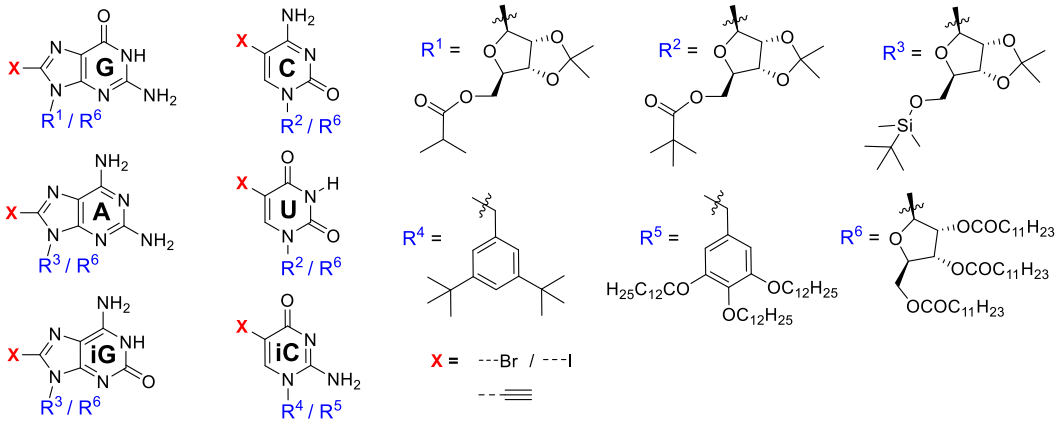


Figure 1.5. General structure of the desired pyrimidine (C, U, iC) and purine (G, A, iG) derivatives.

In general, as mentioned before, the ribose unit was functionalized with bulky lipophilic groups in order to avoid undesired aggregation and to provide characteristic signals in ^{1}H NMR.[134] For instance, the 2' and 3' ribose alcohols were protected as an acetonide group,[135] while the 5' position was functionalized as a *tert*butyl ester (**G1**), an *iso*propyl ester (**C1**, **U1**)[136] or a TBDMS group (**iG1**, **A1**).[137] In order to increase solubility in highly apolar solvents, such as toluene, required for some of our purposes as will be explained later, a second set of derivatives were produced for each one of nucleobases where the ribose was substituted with large alkyl chains. The only exception is **iC**, which was prepared from commercial *iso*cytosine and was equipped with either a 3,5-di*tert*butylbenzyl group or a 3,4,5-tris(dodecyloxy)-benzyl group instead of the natural ribose, *via* nucleophilic substitution reaction with the corresponding benzyl bromide.[138] This change is due to the commercial unavailability of the pristine *iso*cytidine nucleoside analogue.

Common reactions for all nucleosides are halogenation and the introduction of the ethynyl group *via* Sonogashira coupling with TMSA and TMS deprotection. Hence, the

[133] J. Camacho-García, C. Montoro-García, A. M. López-Pérez, N. Bilbao, S. Romero-Pérez, D. González-Rodríguez, *Org. Biomol. Chem.* **2015**, *13*, 4506–4513.
[134] a) D. González-Rodríguez, J. L. J. van Dongen, M. Lutz, A. L. Spek, A. P. H. J. Schenning, E. W. Meijer, *Nature Chem.* **2009**, *1*, 151–155; b) D. González-Rodríguez, P. G. A. Janssen, R. Martín-Rapún, I. De Cat, S. De Feyter, A. P. H. J. Schenning, E. W. Meijer, *J. Am. Chem. Soc.* **2010**, *132*, 4710–4719.
[135] a) B. Zhang, Z. Cui, L. Sun, *Org. Lett.* **2001**, *3*, 275–278; b) Y. Xu, H. Jin, Z. Yang, L. Zhang, *Tetrahedron*, **2009**, *65*, 5228–5239.
[136] I. Manet, L. Francini, S. Masiero, S. Pieraccini, G. P. Spada, G. Gottarelli, *Helv. Chim. Acta*, **2001**, *84*, 2096–2107.
[137] a) S. L. Forman, J. C. Fettinger, S. Pieraccini, G. Gottarelli, J. T. Davis, *J. Am. Chem. Soc.* **2000**, *122*, 4060–4067; b) M. S. Kaucher, J. T. Davis, *Tetrahedron Lett.* **2006**, *47*, 6381–6384.
[138] A. Holý, J. Günter, H. Dvoráková, M. Masojídková, G. Andrei, R. Snoeck, J. Balzarini, E. De Clercq, *J. Med. Chem.* **1999**, *42*, 2064–2086.

variability in the nucleobase compounds is only present at the level of the substituents at N-1 (pyrimidines) or N-7 (purines). For these reasons, the ideal, most convergent synthetic pathway would comprise halogenation, Sonogashira cross-coupling,[2a,3a] followed by TMS deprotection, and finally the modification of the ribose moiety. Using this sequence, a lower number of products would need to be synthesized. This synthetic pathway is preferred and was always prioritized. However, as will be explained below, we deviated from this optimal route whenever it was required because of different synthetic problems, solubility issues, a difficulty to isolate the desired compound, or simply because it was more practical. It was our aim to obtain high amounts (in the multigram scale) of any ethynyl-substituted nucleoside precursor through the most straightforward and convenient route, so we chose the synthetic paths that afforded the highest yields and the easiest purification protocols at some specific steps. For this reason, and as discussed below, the halogenation reaction needed to be performed at different stages and with different methods for each nucleobase. Pyrimidines were selectively iodinated at C-5[139] whereas purines had to be brominated at C-8.[140] The incorporation of iodine was preferred instead of bromine due to the higher reactivity of iodoarenes in metal-catalyzed cross-coupling reactions. Then, the Sonogashira coupling with TMSA and the final deprotection with TBAF was carried out successfully, usually as the last step in the synthetic routes.

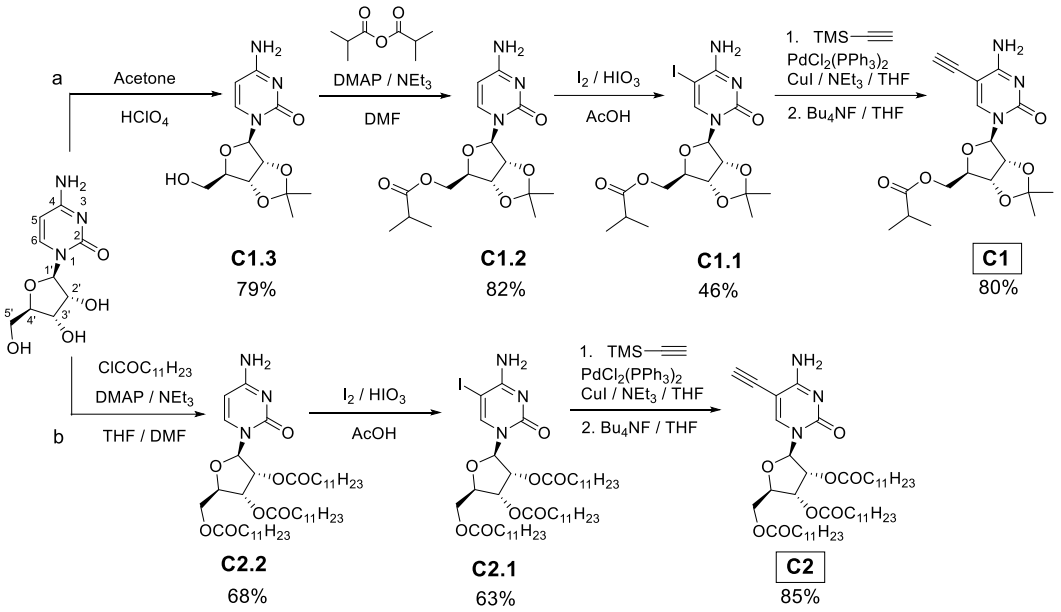
Synthesis of Pyrimidines.

A total of six 5-iodopyrimidines and 5-ethynylpyrimidines were synthesized.

Cytidine (Scheme 1.6). The route to obtain **C1** and **C2** starts directly from commercial cytidine. On one hand (Scheme 1.6a), **C1.3** was obtained according to a previously reported procedure.[13a] Subsequently, the 5' position of **C1.3** was functionalized in the presence of one equivalent of an isobutyric anhydride afford **C1.2** in acceptable yields. This reaction led as well to a subproduct that is doubly substituted on N(4)- and O(5')-positions, and to the starting material because of the -NH_2 group is also reactive in these conditions. On the other hand (Scheme 1.6b), **C2.2** derivative was synthesized by complete acylation of the 2', 3' and 5' ribose alcohols with lauroyl chloride from commercial cytidine. Once the ribose functionalization was completed at all alcohol groups, both derivatives (**C1.2** and **C2.2**) were subjected to a iodination reaction in presence of I_2/HIO_3 in acetic acid, which yielded compounds **C1.1** and **C2.1**. Attempts to iodinate in the first or second step in the first route afforded products that we difficult to purify. Finally, the halogenated bases were subjected to a Sonogashira coupling procedure with TMSA and then treated with TBAF in order to obtain **C1** and **C2** in good yields.

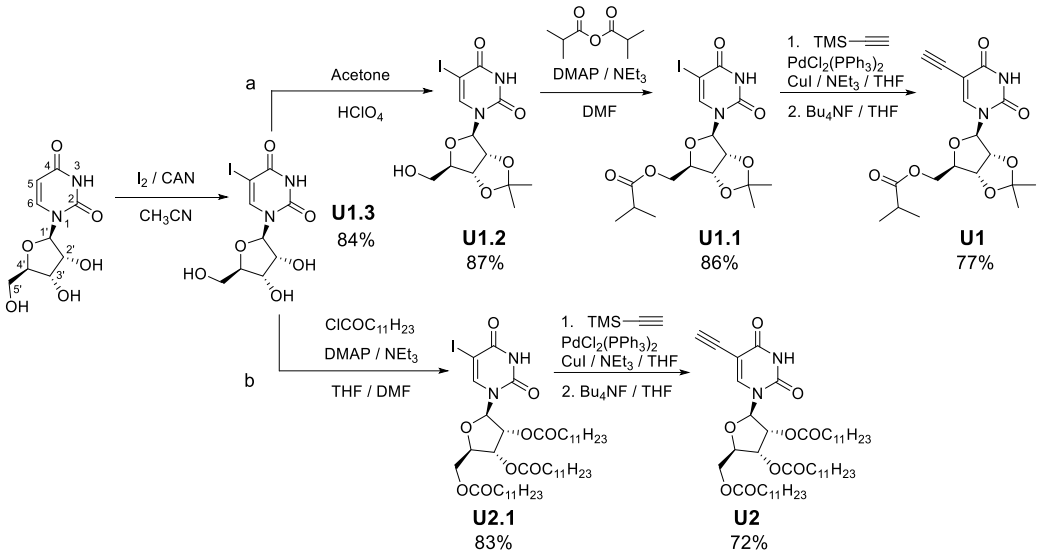
[139] a) M. Bobek, I. Kavai, R. A. Sharma, S. Grill, G. Dutschman, Y.-C. Cheng. *J. Med. Chem.* **1987**, *30*, 2154–2157; b) J. Asakura, M. J. Robins, *J. Org. Chem.* **1990**, *55*, 4928–4933. c) A. Mayer, A. Häberli, C. Leumann, *Org. Biomol. Chem.* **2005**, *3*, 1653–1658.

[140] M. S. Amer, A. M. Amer, A. F. S. Ahmed, W. M. Farouk, *Indian J. Chem., Sect. B*, **2001**, *40B*, 382–385.



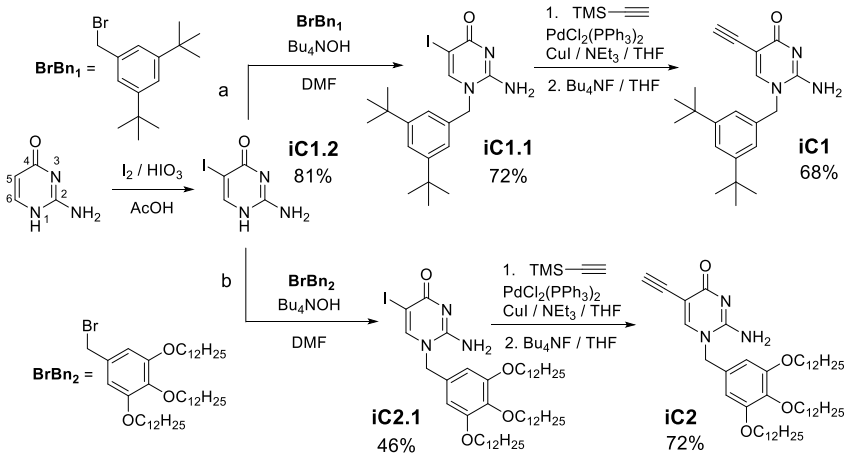
Scheme 1.6. Optimized synthetic routes to cytidines **C1** and **C2**.

Uridine (Scheme 1.7). In contrast to the cytidine synthetic pathway, iodination was the first step used for uridines because the **U1.3** product could be efficiently isolated by straightforward filtration. As shown in Scheme 1.7a, the 2' and 3' ribose alcohols were protected as an acetonide group in the presence of acetone and perchloric acid according to the literature procedure in order to achieve **U1.2**.[13b] Then, a protection of the 5' ribose alcohol position of **U1.2** as an isopropyl ester was performed, obtaining **U1.1**. On the other hand, Scheme 1.7b shows that the complete protection of the ribose alcohols with lauroyl chloride in the presence of 4-dimethylaminopyridine (DMAP) and NEt_3 afforded **U2.1** from the iodinated uridine **U1.3**. Finally, once **U1.1** and **U2.1** was obtained, Sonogashira reaction in the usual conditions with TMSA and subsequent cleavage of the alkyne TMS group affords **U1** and **U2**. Traces of deiodinated products were detected in both cases in this last Sonogashira reaction.



Scheme 1.7. Optimized synthetic route to uridines **U1** and **U2**.

***Iso*cytosine** (Scheme 1.8). Since *iso*cytidine nucleoside is not commercially available, the synthetic pathway toward **iC1** and **iC2** started in this case from *iso*cytosine. As in the route to uridine, the iodination could be carried out as the first step because the product **iC1.2** is efficiently isolated and purified by a simple filtration and washing protocol. Then, 5-iodo*iso*cytosine **iC1.2** was subjected to an alkylation reaction with the respective benzyl bromide in the presence of Bu_4NOH, which proved to be a more efficient base than K_2CO_3 and Cs_2CO_3 because of its solubility in organic solvents, to afford **iC1.1** and **iC2.1**. These reactions led as well to the corresponding N(3)- and O(4)-monoalkylated products in minor amounts. The different isomers were assigned on the basis of their different polarity and with the help of NOESY experiments, as it is explained in the Experimental Section of this Chapter. To conclude, the common Sonogashira coupling-TBAF-mediated TMS cleavage tandem proces was performed on these alkylated-iodo*iso*cytosines, to supply **iC1** and **iC2**.



Scheme 1.8. Optimized synthetic routes to *iso*cytosines **iC1** and **iC2**.

Synthesis of Purines.

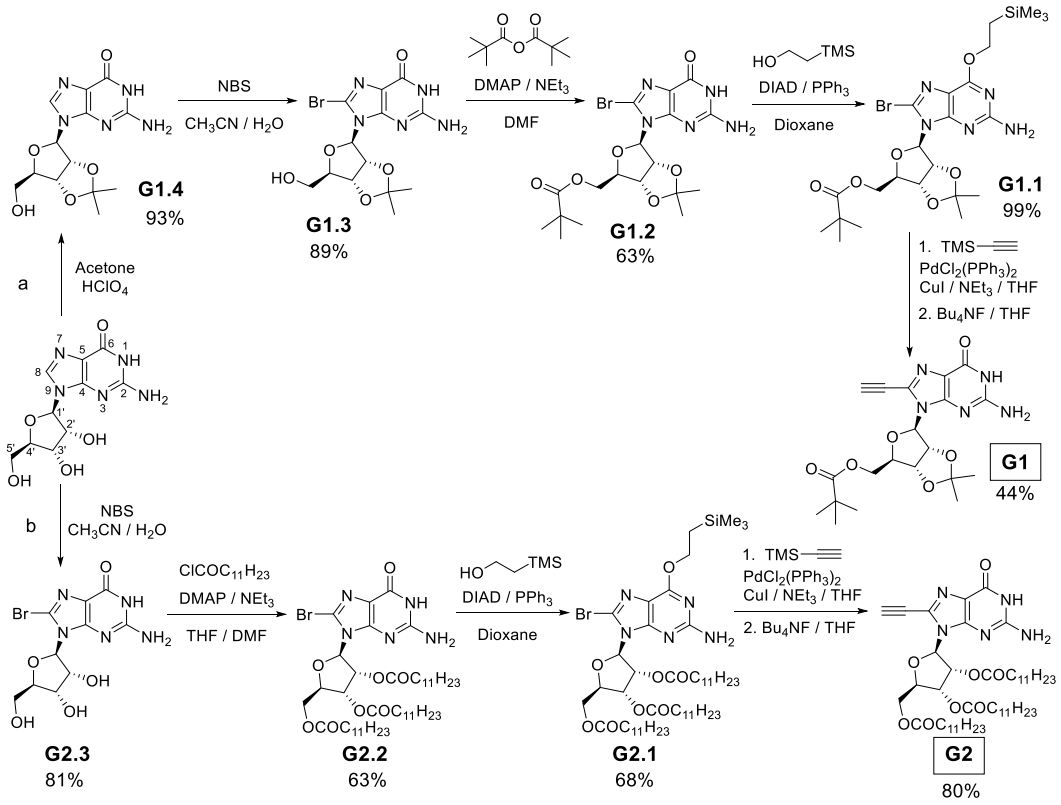
The optimized route to 8-halogenated and 8-ethynylated purines followed a similar synthetic sequence to the one used with the pyrimidines. Purine nucleosides were brominated either in the first step in the route to **G2** and **A2** (or **iG2**) or in the second step in the route to **G1** and **A1**, just after 2',3'-diol protection. The reason is because the C-8 halogenated purine nucleosides are rather sensitive to depurination in the acidic conditions used.[141] Hence, when these two steps were inverted, that is, when guanosine or 2-aminoadenine was instead brominated in first instance and then reacted with acetone in the presence of $HClO_4$ or other acids, the C-*N* glicosidic bond was cleaved quantitatively. Probably, the preparation of 8-iodopurines would have been a better alternative for a more efficient Sonogashira coupling, but their synthesis was found to be problematic for several reasons. First of all, 8-iodopurine nucleosides are even more sensitive to depurination.[19] Second, the electrophilic substitution reaction with iodine is a rather low-yielding reaction that may afford undesired products coming from nucleophilic attack to the activated C-8 position (see below). As with the pyrimidines, the Sonogashira reaction was left as the last step in these synthetic pathways.

Guanosine (Scheme 1.9). The synthetic route to **G1** and **G2** starts from commercial guanosine.

In the case of **G1** (Scheme 1.9a), the 2' and 3' ribose alcohols were first protected as an acetonide group[13] and the aromatic core in **G1.4** was then brominated with NBS[12b] in order to obtain **G1.3**, as previously outlined. Our attempts to iodinate **G1.4** *via* an electrophilic substitution pathway resulted in intramolecular attack of the 5'-OH group to the C-8 position to produce the corresponding cyclic ether. This reaction, as we found out later, had been previously described.[142] On the other hand, all of our efforts to iodinate a guanosine derivative with a fully protected ribose were unsuccessful. After halogenation, the 5' position was functionalized as a *tert*butyl ester to obtain **G1.2**. The yield in this reaction was moderate due to the low chemoselectivity found and the appearance of the doubly acylated derivative on N(2)- and O(5')-positions. In order to try to improve the yield of this last reaction we converted both the N(2) and O(5')-positions into a *tert*butyl ester by reaction with excess anhydride, and then deprotect with NH_3 (7N) in MeOH selectively the N(2) position. These attempts, which were made based on other successful literature results, were unfortunately not successful because of the large amount of starting **G1.3** material recovered instead of the desired O(5')-acylated **G1.2** compound.

For **G2** (Scheme 1.9b), commercial guanosine bromination in the same conditions as those used before (NBS in a mixture of MeCN and H_2O) was the first step, yielding **G2.3** after filtration and washing. Then, complete acylation of the ribose alcohols with lauroyl chloride was carried out in order to synthesize **G2.2**.

[141] P. Lang, C. Gerez, D. Tritsch, M. Fontecave, J.-F. Biellmann, A. Burger, *Tetrahedron*, **2003**, 7315–7322.
[142] Q. Gui-Rong, R. Bo, N. Hong-Ying, M. Zhi-Jie, G. Hai-Ming, *J. Org. Chem.* **2008**, *73*, 2450–2453.

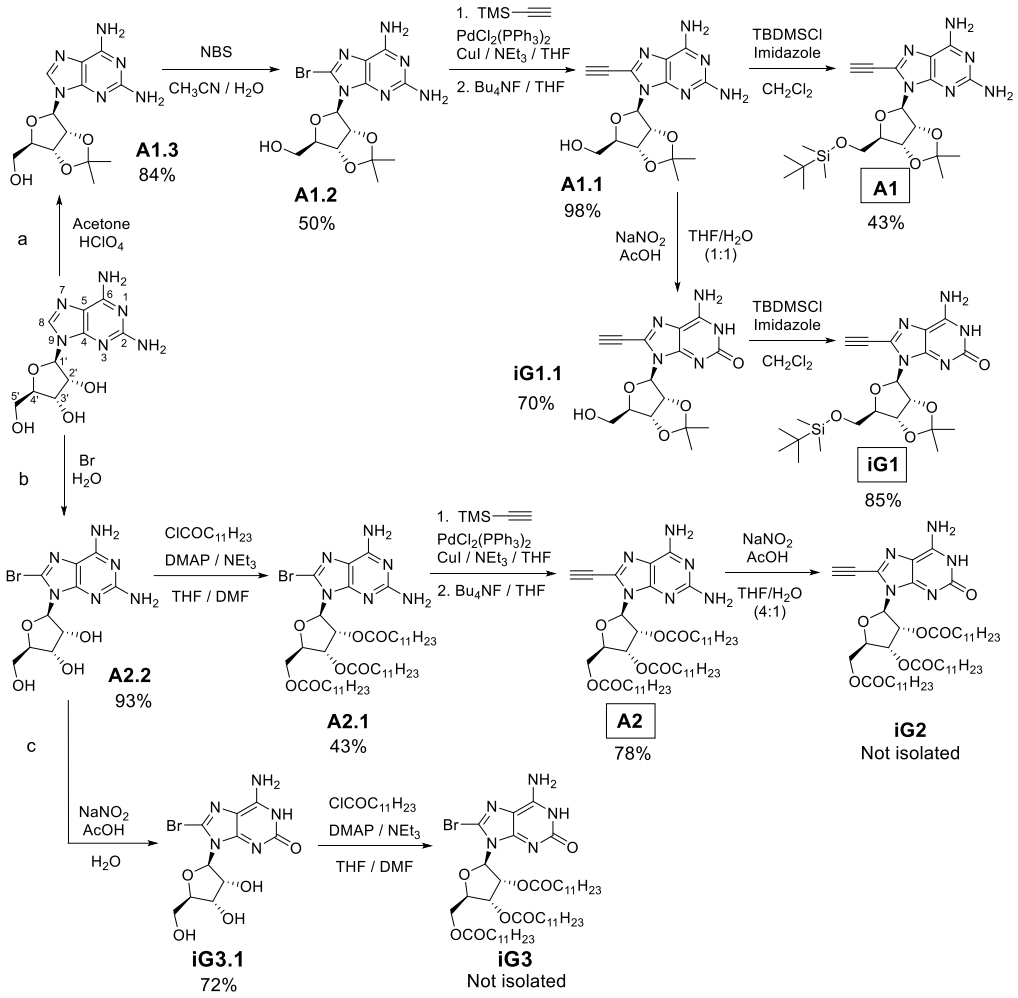


Scheme 1.9. Optimized synthetic route to guanosines **G1** and **G2**.

However, all of our attempts to perform a Sonogashira reaction on the brominated and ribose-protected lipophilic nucleoside derivatives **G1.2** and **G2.2** were unsuccessful, and the starting material was recovered instead. There are in fact many examples in the literature that report on the low reactivity of the guanine heterocycle in transition metal-mediated oxidative addition processes.[143] The low oxidation potential of this base or its ability to coordinate organometallic catalytic species are cited among the causes that would explain such lack of reactivity. The best solution in our hands was the protection of the G-carbonyl group in **G1.2** and **G2.2** as a trimethylsilylethoxy group,[144] in the presence of DIAD and PPh_3, to yield **G1.1** and **G2.1**, compounds that were now active in Pd-catalyzed couplings. A Sonogashira reaction with TMSA and the following removal of the two fluoride-labile protecting groups was achieved in the presence of TBAF in THF, producing **G1** and **G2**.

2-Aminoadenosine (Scheme 1.10). Two 2-aminoadenosine and one *iso*guanosine derivatives were prepared from commercial 2-aminoadenosine (also named 2,6-diaminopurine). As shown in Scheme 1.10, the synthetic pathway is similar to the one optimized for the guanosine derivatives.

[143] E. C. Western, K. H. Shaughnessy, *J. Org. Chem.* **2005**, *70*, 6378–6388.
[144] A. Dumas, N. W. Luedtke. *Chem. Eur. J.* **2012**, *18*, 245–254.



Scheme 1.10. Optimized synthetic route to 2-aminoadenosines **A1** and **A2** and *iso*guanosines **iG1** and **iG2**.

Scheme 1.10 a leads to **A1** and starts with the protection of the 2' and 3' alcohols as an acetonide group and the bromination on the C-8 position in order to obtain sequentially **A1.3** and **A1.2**. Then, a Sonogashira reaction was carried out with TMSA in order to reach **A1.1** in the common reaction conditions. It is interesting to note, in this case, the 2-aminoadenosine purine substrate is reactive in metal-catalyzed couplings. The protection of the 5'-alcohol with TBDMSCl could be therefore left for the last step, resulting in a more convergent synthetic route and generating the desired **A1** product. Another reason why these last steps (Sonogashira and 5'-OH protection) were exchanged in comparison with other synthetic routes was to simplify the isolation of **iG1.1,** as will be explained below).

Scheme 10b leads to the highly lipophilic **A2** nucleoside. Starting from the commercial product, **A2.2** was prepared by a simple bromination reaction with Br_2 in H_2O.[145] Once **A2.2** was obtained, reaction of the ribose alcohols with lauroyl chloride generated **A2.1**. Once again, the aromatic amine groups competed for acylation with the ribose alcohols. Attempts

[145] D. Baranowski, B. Golankiewicz, W. Folkman, M. Popenda, *Nucleos. Nucleot. Nucl. Ac.* **2012**, *30*, 707–719.

to selectively protect these amino groups to enhance O-alkylation as imine groups for example, were fruitless. Finally, in order to synthesize **A2**, a Sonogashira cross coupling was carried out in good yield.

***Iso*guanosine** (Scheme 1.10). The non-natural *iso*guanosine nucleoside can be obtained from 2,6-diaminopurine in a single step by selective hydrolysis at C-2 *via* a diazonium intermediate.[146] For the sake of simplicity, we wanted to perform this process at the end of the reaction route, that is, at the level of **A1** or **A2** but, unfortunately, this could not be done due to several reasons noted below.

In the case of **iG1**, the TBDMS group was found to be sensitive under the acidic conditions employed in the 2-aminoadenosine to *iso*guanosine transformation. That is another reason why, as mentioned before, the final **A1** route presented the Sonogashira coupling before 5'-alcohol protection. **A1.1** is, in contrast, a well-suited substrate for this synthetic transformation. Therefore, once the alkyne-TMS group is cleaved with Bu_4NF, the **A1.1** product is either reacted with TBDMSCl to yield **A1** or with $NaNO_2$ to transform it to **iG1.1**. The latter product can then be functionalized at the 5'-position in the presence of TBDMSCl in similar conditions to those used for **A1.1**, a reaction that affords **iG1.**

In the other synthetic route, unfortunately, all attempts to reach **iG2** derivative were unsuccessful presumably due to problems associated with a difficult homogeneization process rather than to reactivity problems. We believe that the key issue is to find the correct solvent, in this case a THF/H_2O mixture, that is able to dissolve, at least to a minimum extent, both the $NaNO_2$ salt and the strongly lipophilic **A2** derivative. In the best case, we could isolate a single compound with a characteristic amide ^1H NMR signal at 11 ppm characteristic of **iG** compounds in DMSO. However, MS did not reveal the presence of the desired compound. We tried another synthetic route that started from **A2.2** in order to generate **iG3.1** *via* a diazonium intermediate. Unfortunately, the following protection of the ribose alcohols with lauroyl chloride generated a complex mixture of derivatives that hampered chromatographic purification. Finally, after these two attempts to synthesize **iG2**, we decided to conclude with this synthetic route that it was destined to form symmetrically substituted monomers in ABAB systems (see *Chapter 6*). In such Chapter, dedicated to ABAB systems, we will explain other reasons that led us to abandon the synthesis of this compound.

[146] a) S. C. Jurczyk, J. T. Kodra, J.-H. Park, S. A. Benner, T. R. Battersby, *Helv. Quim. Acta,* **1999**, *82*, 1005–1015; b) X. Shi, J. C. Fettinger, M.Cai, J. T. Davis, *Angew. Chem. Int. Ed.* **2000**, *39*, 3124–3127.

1.3. Conclusions.

This Chapter is devoted to synthesize the targeted molecules which form part of the final monomers through the most convergent synthetic route possible. These molecules comprise five π-conjugated central blocks and eleven nucleobases.

Five different π-conjugated molecules formed by aromatic rings, ethynyl linkers and alkyl chains have been synthesized to complete a useful collection of seven central blocks to be used for different purposes in this Thesis, as will be detailed in each of the subsequent Chapters. These building blocks have been designed with the aim to afford the necessary rigidity (π-conjugated body) and solubility (lateral alkyl chains) to the target dinucleoside monomers that can associate in cyclic species with a variable size.

On the other hand, a series of lipophilic nucleobases have been prepared, comprising natural and non-natural derivatives, which are substituted at the 5- (pyrimidines) or 8-position (purines) with either a halogen atom or a terminal triple bond. These include cytidine, *iso*cytosine and uridine as pyrimidine derivatives and guanosine, *iso*guanosine and 2-aminoadenosine as a complementary purine bases. The ribose moiety has been equipped with different bulky groups to afford solubility in the most commonly employed organic solvents and prevent π-π interactions. In another set of derivatives, the ribose groups have been functionalized with long alkyl chains in order to improve further the solubility in more apolar organic solvents, such as toluene or carbon tetrachloride. The synthetic pathway leading to the final ethynylated compounds have been optimized for each nucleobase attending to reactivity problems, convenience, ease of purification, and overall yields. The final results indicate that the choice of the synthetic route is not trivial, and that each nucleobase requires a particular optimized protocol to reach the final halogenated/ethynylated products.

We have thus achieved the preparation of a new collection of “supramolecular synthons” that will be used for the purposes of this Thesis as will be detailed in each Chapter. For that, the synthesis of symmetrically substituted and unsymmetrically substituted monomers bearing complementary and non-complementary nucleobases can guide self-assembly to form a tetramer in solution constituted by 1, 2, 3 or 4 components, depending of the design chosen.

1.4. Experimental Section.

1.4.1. General Methods.

Chemicals were purchased from commercial suppliers and used without further purification. Solid, hygroscopic reagents were dried in a vacuum oven before use. *N*-Bromosuccinimide (NBS) was recrystallized from water. Reaction solvents were thoroughly dried before use employing standard methods or using a solvent purification system Innovative Technology Inc. MD-4-PS (CH_3CN, THF, CH_2Cl_2, Et_2O). Furthermore, THF destined to form part of a Sonogashira reaction through a deoxygenation by three freeze-pump-thaw cycles with argon was dried with Na (drier) and benzophenone (indicator).

Chromatography. Reactions were monitored by *Thin Layer Chromatography* (TLC) using 0.2 mm aluminium sheets precoated with silica gel 60 F254 (Merck). TLC plates were inspected with a UV lamp featuring both long-wavelenght UV light (365 nm) and short-wavelenght UV-light (254 nm).

Column chromatography was carried out on silica gel *Merck-60* (230-400 mesh, 60 Å). Eluent relative volume/volume ratio are indicated in each case.

Nuclear Magnetic Resonance Spectroscopy (NMR)**:** ^{1}H NMR, ^{13}C NMR and ^{19}F NMR spectra were recorded with a *BRUKER AC-300* (300 MHz) instrument at the Organic Chemistry department or a *BRUKER XRD-500* (500 MHz) instrument at the Interdepartmental Service for Research (SIdI) at the UAM. The temperature was actively controlled at 298 K. Chemical shifts (δ) are measured in parts-*per*-million (ppm) using the signals of the deuterated solvent as the internal standard [$CDCl_3$, calibrated at 7.26 ppm (^{1}H) and 77.0 ppm (^{13}C); DMSO-d_6 (2.50 and 39.5 ppm); toluene-d_8 (7.09 and 137.9 ppm); THF-d_8 (1.73, 3.58 and 25.5 ppm); DMF-d_7 (8.03, 2.75 and 163.15, 34.89 ppm); $CDCl_2CDCl_2$ (6.00 and 73.78 ppm)]. The used deuterated solvents are indicated in each case. The following abbreviations have been used for the spectra description: s (singlet), d (doublet), t (triplet), q (quartet), quint (quintet), sext (sextet), m (multiplet), dd (doublet of doublets), td (triplet of doublets), bs (broad signal).

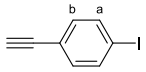
Mass Spectrometry (MS) and **High Resolution-Mass Spectrometry** (HRMS) spectra were measured at the SIdI on a *VG AutoSpec* apparatus (FAB) or an *Applied Biosystems QSTAR* equipment (ESI) in the positive or negative modes. MALDI-TOF-MS spectra were obtained from a BRUKER ULTRAFLEX III instrument equipped with a nitrogen laser operating at 337 nm. The matrix employed is indicated in each case.

1.4.2. Synthesis and Characterization.

The synthesis and characterization of compounds 4,4'-Diiodo-2,2',6,6'-tetramethylbiphenyl (**B2''**),[147] 1,4-diiodo-2,5-dimethoxybenzene (**B3.4**),[131a] 2,5-diiodo-1,4-hydroquinine (**B3.3**),[131b] **B3**,[130] **C1.3**,[135a] **U1.3**,[135b] **iC1.2**,[139c] **BrBn$_2$**,[148] **G1.4**,[135] **G1.3**,[134b] **G2.3**,[135] **A1.3**,[135a] **A2.2**,[145] **iG3.1**,[146] **G$_{alk2}$**, **G$_{alk10}$**, and **A$_{alk10}$**[149]. The ^{1}H and ^{13}C NMR spectra of most of the compounds described below can be found in the Supporting Information of ours papers: *Org. Biomol. Chem.* **2015**, *13*, 4506–4513 (ref. 133) and, *Eur. J. Org. Chem*. **2015**, *32*, 7160–7175 (ref. 149).

Synthesis of the dihalogenated blocks.

- ***Standard Procedure A*** for the Sonogashira coupling with TMSA and subsequent alkyne-TMS group deprotection. A dry THF/NEt_3 (4:1) solvent mixture was subjected to deoxygenation by three freeze-pump-thaw cycles with argon. Then, this solvent was added over the system containing either the corresponding halogenated base (1 eq), or the corresponding dihalogenated central block (indicated in each case) CuI (0.01 eq) and $PdCl_2(PPh_3)_2$ (0.02 eq). The mixture was stirred at room temperature during a few minutes. Then, trimethylsilylacetylene (TMSA; eq indicated in each case) was added dropwise. The reaction was stirred under argon at a given temperature for a period of time (indicated in each case) until completion, which was monitored by TLC. Then, the mixture was filtered over a celite plug and the solvent evaporated under vacuum. The resulting crude was placed in a round-bottomed flask equipped with a magnetic stirrer, THF was added and the mixture was stirred at room temperature until the solid was dissolved. Then, hydrated tetrabutylammonium fluoride (TBAF·$3H_2O$; eq indicated in each case) was added at 0 ^oC, and the mixture was stirred until reaction completion, which was monitored by TLC (approximately 1 hour in all cases). The solvent was evaporated at reduced pressure and the product was purified by column chromatography (eluent indicated in each case). The resulting solid was finally reprecipitated with cold acetonitrile.



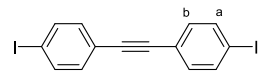
B2.1. Following *Standard Procedure A,* to a solution of commercial iodoarene **B1** (30.31 mmol, 5 g), $Pd(PPh_3)_2Cl_2$ (0.12 mmol, 85 mg) and CuI (0.06 mmol, 11 mg) in THF/NEt_3 (4:1) (80 mL), TMSA (1 eq, 6 mmol, 0.929 mL) was added and the mixture was stirred at 40 ^oC overnight. After reaction completion, the deprotection reaction is carried out without further purification following the *Standard Procedure A*. The crude product obtained after solvent evaporation was suspended in dry THF (100 mL) and TBAF·$3H_2O$ (1.2 eq, 7.2 mmol, 2.27 g) was added. After reaction completion the solvent was evaporated and the resulting residue was purified by column chromatography using cyclohexane as eluent. 0.89 g of compound **B2.1** was obtained as a white solid (65% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.54 (d, J = 8.4 Hz, 2H, H^a), 7.09 (d, J = 8.4 Hz, 2H, H^b), 2.98 (s, 1H, -CCH).

- ***Standard Procedure B*** for the Sonogashira coupling with halogenated central block and subsequent ethynyl-central block. A dry THF/NEt_3 (4:1) solvent mixture was subjected to deoxygenation by three freeze-pump-thaw cycles with argon. Then, this solvent was added over the system containing the corresponding halogenated central block (1 eq), CuI (0.01 eq) and $PdCl_2(PPh_3)_2$ (0.02 eq). The mixture was stirred at room temperature during a few minutes. Then, the substituted block (eq indicated in each case) was added dropwise. The reaction was stirred under argon at a given temperature for a period of time (indicated in each case) until completion, which was monitored by TLC. The order between halogenated and ethynylated central blocks was changed in order to favour the mono- or doubly substituted monomer. Then, the mixture was filtered over a celite plug and the solvent evaporated under vacuum. The resulting crude (if required) was placed in a round-bottomed flask equipped with a magnetic stirrer, THF was added and the mixture was stirred at room temperature until the solid was dissolved. Then, hydrated tetrabutylammonium fluoride (TBAF·$3H_2O$; eq indicated in each case) was added at 0 ^oC, and the mixture was stirred until reaction completion, which was monitored by TLC (approximately 1 hour in all cases). The solvent was evaporated at

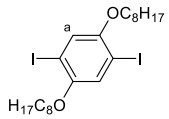
[147] D. Vonlanthen, J. Rotzler, M. Neuburger, M. Mayor, *European Journal of Organic Chemistry,* **2010**, 120–133.
[148] L. Lee, Y. Zao, *J. Am. Chem. Soc.* **2014**, *136*, 5579–5582.
[149] N. Bilbao, V. Vázquez-González, M. T. Aranda, D. González-Rodríguez, *Eur. J. Org. Chem.* **2015**, *54*, 7160–7175.

reduced pressure and the product was purified by column chromatography (eluent indicated in each case). The resulting solid was finally reprecipitated with cold acetonitrile.



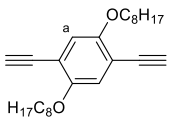
B2. Following *Standard Procedure B,* to a solution of iodoarene (19.5 mmol, 6.41 g), $PdCl_2(PPh_3)_2$ (78 µmol, 55 mg) and CuI (40 µmol, 7.6 mg) in THF/NEt_3 (4:1) (50 mL) was added dropwise **B2.1** (3.9 mmol, 0.89 g). Then, the mixture was stirred at 40 ^oC overnight. After reaction completion the solvent was evaporated and the resulting residue was purified by column chromatography using cyclohexane as eluent. 1.31 g of compound **B2** was obtained as a white solid (78% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.69 (d, J = 8.4 Hz, 4H, H^b), 7.24 (d, J = 8.5 Hz, 4H, H^a).



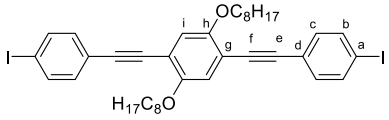
B3.2.[130] To a solution of **B3.3** (55.26 mmol, 20 g) and KOH (166 mmol, 9.3 g) in dry DMF at 0 ^oC (200 mL) was added dropwise 1-bromooctane (121 mmol, 13.05 g) and stirred overnight. After reaction completion the solvent was evaporated and the resulting residue was washed three times with iPr_2O, then with $HCl_{(aq)}$ 0.1 M and finally dried with $MgSO_4$. The resulting crude was purified by column chromatography using cyclohexane/DCM (10:1) as eluents. **B3.2** was obtained as a white solid (78% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.21 (s, 2H, H^a), 3.97 (t, J = 6.4 Hz, 4H, OCH_2-), 1.84 (p, J = 6.4 Hz, OCH_2CH_2-), 1.63 – 1.22 (m, 20H,-CH_2-), 0.92 (d, J = 6.8 Hz, 6H, -CH_3).



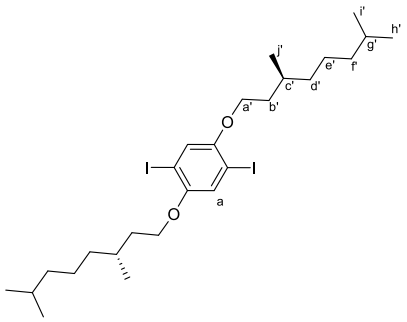
B3.1.[130] Following *Standard Procedure A,* to a solution of **B3.2** (5.86 mmol, 3.4 g), $PdCl_2(PPh_3)_2$ (0.12 mmol, 82 mg) and CuI (0.06 mmol, 11 mg) in THF/NEt_3 (4:1) (50 mL) TMSA (4 eq, 23.4 mmol, 3.6 mL) was added and the mixture was stirred at 40 ^oC overnight. After reaction completion, the deprotection reaction is carried out without further purification following the *Standard Procedure A*. The crude product obtained after solvent evaporation was suspended in dry THF (50 mL) and TBAF·$3H_2O$ (2.4 eq, 14 mmol, 4.43 g) was added. After reaction completion the solvent was evaporated and the resulting residue was purified by column chromatography on silica gel eluted with cyclohexane /DCM (8:1). 2.7 g of compound **B3.1** was obtained as a white solid (86% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 6.99 (s, 2H, H^a), 4.01 (t, J = 6.6 Hz, 4H, OCH_2-), 3.37 (s, 2H, -CH), 1.93 – 1.76 (m, 4H, OCH_2CH_2-), 1.63 – 1.22 (m, 20H, -CH_2-), 1.01 – 0.85 (m, 6H, -CH_3).



B3.[130] Following *Standard Procedure B,* to a solution of iodoarene (9.15 mmol, 3.01 g), $PdCl_2(PPh_3)_2$ (7 µmol, 5 mg) and CuI (4 µmol, 0.7 mg) in THF/NEt_3 (4:1) (50 mL), **B3.1** (0.19 mmol, 70 mg) was added dropwise and the mixture was stirred at 40 ^oC overnight. After reaction completion, the resulting residue was purified by column chromatography on silica gel eluted with cyclohexane /toluene (10:1). 107 mg of compound **B3** was obtained as a yellow solid (75% yield).

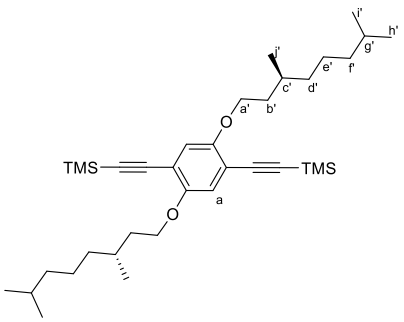
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.62 (d, J = 8.4 Hz, 4H, H^b), 7.18 (d, J = 8.3 Hz, 4H, H^c), 6.92 (s, 2H, H^i), 3.95 (t, J = 6.4 Hz, 4H, OCH_2-), 1.84 – 1.69 (m, 4H, OCH_2CH_2-), 1.52 – 1.11 (m, 20H, -CH_2-), 0.89 – 0.73 (m, 6H, -CH_3).



B4.5. To a solution of **B3.3** (7.53 mmol, 2.7 g) and KOH (22.3 mmol, 1.3 g) in dry DMF at 0 ^oC (200 mL) was added dropwise **Br-R*** (22.6. mmol, 4.98 g) and stirred overnight. After reaction completion the solvent was evaporated and the resulting residue was extracted three times with iPr_2O, then with $HCl_{(aq)}$ 0.1 M and finally dried with $MgSO_4$. The resulting crude was purified by column chromatography using toluene/AcOEt (10:1) as eluents. Compound **B4.5** was obtained as a white solid (81% yield).

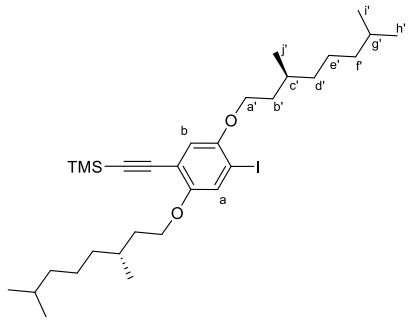
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.17 (s, 2H, H^a), 3.96 – 3.65(m, 4H, OCH_2), 1.98 – 1.82(m. 2H, C$H^{c'}$), 1.78 – 1.62 (m, 4H, C$H^{2b'}$), 1.56 – 1.41 (m, 2H, C$H^{g'}(CH_3)_2$), 1.34 – 1.18 (m, 12H, C$H_2^{d',e',f'}$), 0.95 (d, 6H, C$H_3^{j'}$), 0.87 (d, 12H, C$H_3^{i',h'}$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 152.9, 122.7, 86.3, 68.7, 39.2, 37.2, 36.1, 29.8, 28.0, 24.7, 22.8, 22.7, 19.7.



B4.3. Following *Standard Procedure A,* to a solution of **B4.5** (6.23 mmol, 4 g), $PdCl_2(PPh_3)_2$ (0.12 mmol, 87 mg) and CuI (0.06 mmol, 12 mg) in THF/NEt_3 (4:1) (60 mL) TMSA (1.5 eq, 9.34 mmol, 1.43 mL) was added and the mixture was stirred at 40 ^oC overnight. After reaction completion the solvent was evaporated and the resulting residue was purified by column chromatography on silica gel eluted with toluene/AcOEt (10:1). 52% of compound **B4.3** was obtained as a brown oil.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.18 (s, 1H, H^a), 3.95 (td, J = 6.4, J' = 2.5 Hz, 4H, OCH_2-), 1.92 – 1.01 (m, 20H, -$CH_2^{b'\text{-}g'}$-), 0.87 (dd, J = 6.5, J' = 1.6 Hz, 6H, -$CH_3^{j'}$), 0.85 (d, J = 6.7 Hz, 12H, -$CH_3^{i',h'}$), 0.18 (d, J = 2.9 Hz, 18H, -Si(CH_3)).

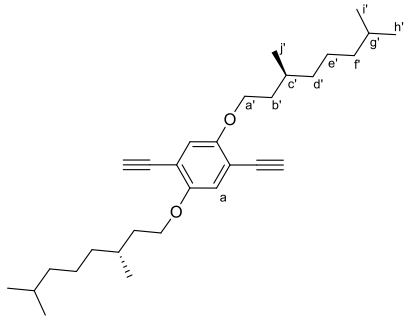


B4.4. Following *Standard Procedure A,* to a solution of **B4.5** (6.23 mmol, 4 g), $PdCl_2(PPh_3)_2$ (0.12 mmol, 87 mg) and CuI (0.06 mmol, 12 mg) in THF/NEt_3 (4:1) (60 mL) TMSA (1.5 eq, 9.34 mmol, 1.43 mL) was added and the mixture was stirred at 40 ^oC overnight. After reaction completion the solvent was evaporated and the resulting residue was purified by column chromatography on silica gel eluted with toluene/AcOEt (10:1). 22% of compound **B4.4** was obtained as a brown oil.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.23 (s, 1H, H^a), 6.79 (s, 1H, H^b), 3.93 (td, J = 6.4, 2.5 Hz, 4H, OCH_2-), 1.90 – 0.99 (m, 20H, -$CH_2^{b'\text{-}g'}$-), 0.90 (dd, J = 6.5, J' = 1.6 Hz, 6H, -$CH_3^{j'}$), 0.82 (d, J = 6.7 Hz, 12H, -$CH_3^{i',h'}$), 0.21 (d, J = 2.9 Hz, 9H, -Si(CH_3)).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) 154.1, 117.8, 113.8, 82.4, 79.5, 68.3, 39.6, 37.1, 36.5, 29.6, 28.5, 26.7, 24.7, 22.7, 22.6, 19.8, 3.5.

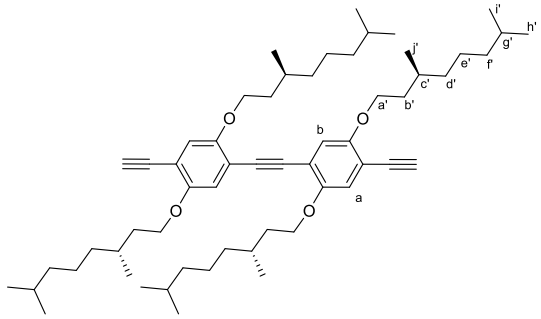
HRMS (ESI+): Calculated for $C_{31}H_{53}IO_2Si$: 613.2860 $[M+H]^+$. Found: 635.2752 $[M+Na]^+$.



B4.2. To a solution of compound **B4.3** (6.86 mmol, 4 g) in THF (50 mL) and TBAF·$3H_2O$ (2.4 eq, 16 mmol, 5.19 g) was added. After reaction completion the solvent was evaporated and the resulting residue was purified by column chromatography on silica gel eluted with toluene/AcOEt (10:1). 2.7 g of compound **B3.1** was obtained as a brown oil (89% yield).

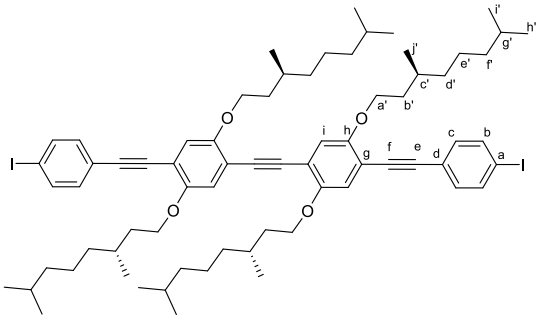
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 6.95 (s, 2H, H^a), 3.99 (ddt, J = 9.3, J' = 6.2, J'' = 2.8 Hz, 4H, OCH_2), 3.32 (s, 1H, CCH), 1.98 – 1.01 (m, 20H, -$CH_2^{b'\text{-}g'}$-), 0.94 (d, J = 6.4 Hz, 6H, -$CH_3^{j'}$), 0.87 (d, J = 6.6 Hz, 12H, -$CH_3^{i',h'}$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 154.0, 117.6, 113.2, 82.5, 79.8, 68.0, 39.2, 37.3, 36.1, 29.8, 28.0, 26.9, 24.7, 22.7, 22.6, 19.8.

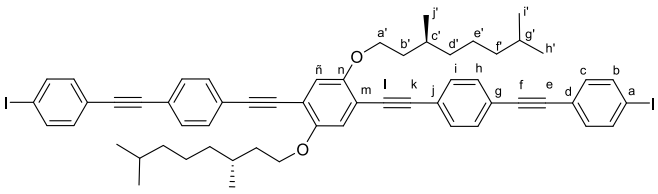


B4.1. Following *Standard Procedure B,* to a solution of **B4.2** (1.6 mmol, 704 mg), $PdCl_2(PPh_3)_2$ (0.05 mmol, 3.75 mg) and CuI (0.03 mmol, 0.51 mg) in THF/NEt_3 (4:1) (15 mL), **B4.4** (1 eq, 0.27 mmol, 164 mg) was added dropwise and the mixture was stirred at 40 ^oC overnight. After reaction completion, the deprotection reaction is carried out without further purification following the *Standard Procedure A*. The crude product obtained after solvent evaporation was suspended in dry THF (50 mL) and TBAF·$3H_2O$ (2.4 eq, 6.43 mmol, 202 mg) was added. After reaction completion the solvent was evaporated and the resulting residue was purified by column chromatography on silica gel eluted with cyclohexane/toluene (8:1). 166 mg of compound **B4.1** was obtained as a brown oil (73% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 6.99 (d, J = 4.2 Hz, 4H, H^a, H^b), 4.03 (q, J = 6.9 Hz, 8H, -OCH_2), 3.34 (s, 2H, -CCH), 1.99 – 1.02 (m, 40H, -C$H_2^{b'-g'}$-), 1.01 – 0.92 (m, 12H, -C$H_3^{j'}$), 0.86 (dd, J = 10.5, J' = 6.5 Hz, 24H, -C$H_3^{i',h'}$).
^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 154.2, 153.4, 128.2, 117.9, 116.9, 114.8, 112.6, 91.3, 82.3, 80.1, 68.2, 68.0, 39.3, 37.3, 36.3, 36.1, 30.1, 29.9, 28.0, 28.0, 24.7, 22.7, 22.7, 22.6, 22.6, 19.9, 19.8.
HRMS (ESI+): Calculated for $C_{58}H_{90}O_4$: 851.6839 $[M+H]^+$. Found: 873.6768 $[M+Na]^+$.

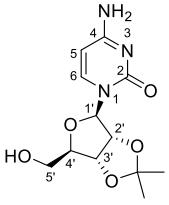


B4. Following *Standard Procedure B,* to a solution of iodoarene (6.77 mmol, 2.23 g), $PdCl_2(PPh_3)_2$ (5 µmol, 3 mg) and CuI (2 µmol, 0.5 mg) in NEt_3/THF (4:1) (30 mL), **B4.1** (0.19 mmol, 70 mg) was added dropwise and the mixture was stirred at 40 ^{o}C overnight. After reaction completion, the resulting residue was purified by column chromatography on silica gel eluted with cyclohexane/toluene (5:1). 188 mg of compound **B4** was obtained as a yellow solid (79% yield).
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.69 (d, J = 8.2 Hz, 4H, H^b), 7.25 (d, J = 8.1 Hz, 4H, H^c), 7.02 (d, J = 4.4 Hz, 4H, H^i), 4.07 (q, J = 7.5 Hz, 8H, -OCH_2), 2.00 – 1.03 (m, 40H, -C$H_2^{b'-g'}$-), 0.98 (dd, J = 6.5, J' = 3.3 Hz, 12H, -C$H_3^{j'}$), 0.85 (dd, J = 7.3, J' = 2.4 Hz, 24H, -C$H_3^{i',h'}$).
^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 153.5, 137.5, 133.0, 123.0, 117.0, 116.9, 114.6, 113.6, 94.1, 93.9, 91.6, 87.6, 68.2, 67.9, 39.3, 37.4, 37.3, 36.4, 36.3, 30.1, 30.0, 28.0, 28.0, 24.8, 24.7, 22.7, 22.6, 19.9, 19.7.
HRMS (ESI+): Calculated for $C_{70}H_{96}I_2O_4$: 1255.5398 $[M+H]^+$. Found: 1255.5392 $[M+H]^+$.

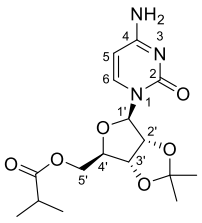


B.5. Following *Standard Procedure B,* to a solution of **B2** (2.32 mmol, 1 g), $PdCl_2(PPh_3)_2$ (9 µmol, 7 mg) and CuI (5 µmol, 0.9 mg) in NEt_3/THF (4:1) (15 mL), **B4.2** (0.23 mmol, 0.1 g), was added dropwise and the mixture was stirred at 40 ^{o}C overnight. After reaction completion, the resulting residue was purified by column chromatography on silica gel eluted with cyclohexane/toluene (8:1). 193 mg of compound **B.5** was obtained as a yellow solid (80% yield).
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.70 (d, J = 8.3 Hz, 4H, H^b), 7.51 (d, J = 3.9 Hz, 8H, $H^{i,h}$), 7.26 (d, J = 8.4 Hz, 4H, H^c), 7.03 (s, 2H, $H^{ñ}$), 4.06 (q, J = 10.0, 8.5 Hz, 4H, -OCH_2), 2.02 – 1.04 (m, 20H, -C$H_2^{b'-g'}$-), 0.99 (d, J = 6.4 Hz, 6H, -C$H_3^{j'}$), 0.86 (d, J = 6.5 Hz, 12H, -C$H_3^{i',h'}$)
^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 153.7, 137.6, 133.8, 133.1, 132.4, 131.5, 123.6, 122.7, 122.6, 116.7, 113.9, 94.7, 94.4, 90.6, 90.4, 88.2, 67.9, 39.3, 37.4, 36.3, 30.0, 28.0, 24.8, 22.8, 22.7, 19.8.
HRMS (ESI+): Calculated for $C_{58}H_{60}I_2O_2$: 1043.2761 $[M+H]^+$. Found: 1066.2689 $[M+H+Na]^+$.

Synthesis of the cytidine derivatives.

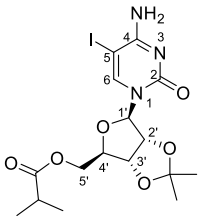


C1.3. C1.3 was synthesized according to a literature procedure.[135a] The pristine cytidine (102.8 mmol, 25 g) was dissolved in acetone (800 mL) and stirred. Then, perchloric acid (70%) (154.2 mmol , 13.29 mL) was added dropwise until completion. After reaction completion, $MgSO_4$ was added and the mixture is further stirred for 1 hour in order to react with the remaining moisture. After that, K_2CO_3 was added and the reaction was stirred at room temperature overnight. The product was then purified by chromatography on silica gel eluted with acetone/$CHCl_3$ (3:1), obtaining a white solid (23 g, 79%).
^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 7.69 (d, J = 8,15 Hz, 1H, H^5), 7.24 (s (broad), 2H, NH_2), 5.77 (s, 1H, $H^{1'}$), 5.70 (d, J = 8,15 Hz, 1H, H^6), 4.98 (d, J = 6.2 Hz, 1H, $H^{2'}$), 4.35 – 4.10 (m, 2H, $H^{3'}$, $H^{4'}$), 4.15 – 4.01 (m, 2H, $H^{5'}$), 1.48 (s, 3H, -OC(CH_3)), 1.29 (s, 3H, -OC(CH_3)).



C1.2. The protected **C1.3**[135a] (20.33 mmol, 5.76 g) and a catalytic amount of DMAP (0.2 eq, 4.07 mmol, 497 mg) were dissolved in 106 mL of dry MeCN. Then NEt_3 (1.5 eq, 30.50 mmol, 4.24 mL) and isobutyric anhydride (1.1 eq, 22.37 mmol, 3.71 mL) were added. The mixture is stirred at room temperature overnight. The reaction is followed by TLC and, once finished, 2 mL MeOH were added and the mixture is further stirred for 30 min in order to react with the remaining anhydride. The solvent was removed and the solid was extracted with water (3 x 100 mL). The organic layer was dried over $MgSO_4$ and the solvent is evaporated. The product was then purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1), obtaining a yellow solid (5.91 g, 82%).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.39 (s (broad), 1H, NH_2), 7.70 (d, J = 8,15 Hz, 1H, H^5), 6.68 (s (broad), 1H, NH_2), 5.85 (s, 1H, $H^{1'}$), 5.70 (d, J = 8,15 Hz, 1H, H^6), 4.78 (d, J = 6.2 Hz, 1H, $H^{2'}$), 4.35 – 4.10 (m, 2H, $H^{3'}$, $H^{4'}$), 4.15 – 4.01 (m, 2H, $H^{5'}$), 2.55 – 2.49 (m, 1H, -COCH), 1.48 (s, 3H, -OC(CH_3)), 1.29 (s, 3H, -OC(CH_3)), 1.11 (s, 3H, -COCH-(CH_3)$_2$), 1.09 (s, 3H, -COCH-(CH_3)$_2$).

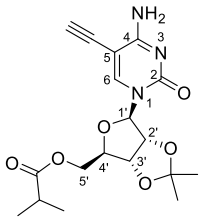


C.1.1.[17a] A suspension of **C1.2** (16.72 mmol, 5.91 g), I_2 (15.0 mmol, 3.78 g) and HIO_3 (27.9 mmol, 4.92 g) in 148 mL acetic acid was stirred overnight at 40 ^{o}C. Once the reaction was completed, the insoluble HIO_3 was filtered and discarded. A 1:1 AcOEt/Et_2O mixture (250 mL) was added to the reaction mixture and this was washed with 3 x 150 mL water, 3 x 150 mL $NaHCO_3$ (sat), 1 x 150 mL $Na_2S_2O_3$ (sat) and finally 1 x 150 mL water. The organic layer was dried over $MgSO_4$ and the solvent was evaporated. The product was collected after recrystallization in a mixture Et_2O/iPr$_2$O as a yellow solid (3.72 g, 46%).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.31 (s, 1H, H^6), 7.92 (s (broad), 1H, NH_2), 5.71 (s, 1H, $H^{1'}$), 5.71 (s, 1H, $H^{2'}$),5.00 (dd, J = 6.4, J' = 1.8 Hz , 1H, $H^{3'}$), 4.26 – 4.16 (m, 3H, $H^{4'}$, $H^{5'}$), 2.60 – 2.53 (m, 1H, -COCH), 1.47 (s, 3H, -OC(CH_3)), 1.28 (s, 3H, -OC(CH_3)), 1.10 (s, 3H, -COCH-(CH_3)$_2$), 1.07 (s, 3H, -COCH-(CH_3)$_2$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 176.6, 164.3, 154.6, 148.3, 114.2, 95.8, 85.9, 85.6, 81.2, 64.3, 56.9, 34.0, 27.2, 25.4, 19.2, 19.1.

HRMS (ESI+)**:** Calculated for $C_{16}H_{22}IN_3O_6$: 480.0553 $[M+H]^+$. Found: 480.0625 $[M+H]^+$.

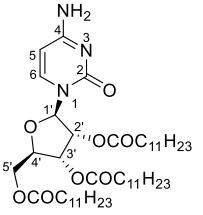


C1. **C1** was prepared following the *Standard Procedure A* for a Sonogashira coupling. **C1.1** (2.08 mmol, 1.00 g), $PdCl_2(PPh_3)_2$ (0.042 mmol, 29.0 mg) and CuI (0.03 mmol, 4.0 mg) were dissolved in the THF/Et_3N (4:1) mixture (20 mL). Then TMSA (6 mmol, 0.79 g) was added and the mixture was stirred at 40 ^{o}C during 24 h. The product was used in the next reaction without purification. Following *Standard Procedure B*, TBAF·3H_2O (3.2 mmol, 1.0 g) was added over a THF (20 mL) solution of the previous crude product. **C1** was purified by chromatography on silica gel eluted by $CHCl_3$/MeOH; (20:1). A final recrystallization using CH_2Cl_2/cyclohexane yielded **C1** as a brown solid (0.61 g, 80%).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.04 (s, 1H, H^6), 7.86 (s (broad), 1H, NH_2), 6.96 (s (broad), 1H, NH_2), 5.76 (s, 1H, $H^{1'}$), 4.99 (dd, J = 6.4, J' = 1.8 Hz , 1H, $H^{2'}$), 4.80 (dd, J = 6.4, J' = 3.1 Hz, 1H, $H^{3'}$), 4.36 (s, 1H, -CCH), 4.29 – 4.17 (m, 2H, $H^{4'}$, $H^{5'}$), 2.62 – 2.55(m, 1H, -COCH), 1.47 (s, 3H, -OC(CH_3)), 1.28 (s, 3H, -OC(CH_3)), 1.08 (d, J = 7.0 Hz, 6H, -COCH-(CH_3)$_2$).

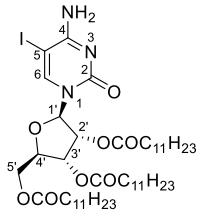
^{13}C-NMR (75 MHz, DMSO-d_6) δ(ppm) = 175.8, 164.5, 153.1, 147.6, 112.9, 93.8, 89.3, 86.1, 84.8, 84.3, 80.9, 75.2, 63.9, 33.1, 26.9, 25.1, 18.7, 18.7, 13.4.

MS (FAB+): 378.2 $[M+H]^+$.



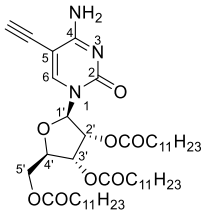
C2.2. The commercial cytidine (4.11 mmol, 1 g) and a catalytic amount of DMAP (0.6 eq, 2.46 mmol, 0.3 g) were dissolved in 30 mL of dry THF/DMF (2:1). Then NEt_3 (4.5 eq, 18.5 mmol, 2.5 mL) and lauroyl chloride (3.1 eq, 12.74 mmol, 2.94 mL) were added. The mixture is stirred at room temperature overnight. The reaction is followed by TLC and, once finished, 5 mL MeOH were added and the mixture is further stirred for 30 min in order to react with the remaining chloride. The solvent was removed and the oil was extracted with water (3 x 20 mL). The organic layer was dried over $MgSO_4$ and the solvent is evaporated. The product was then purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), obtaining **C2.2** as a yellow oil (2.20 g, 68%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.02 (s (broad), 2H,), 7.83 (d, *J* = 7.6 Hz, 1H, H^6), 7.39 (d, *J* = 7.5 Hz, 1H, H^5), 6.08 (d, *J* = 4.3 Hz, 1H, $H^{1'}$), 5.40 – 5.29 (m, 1H, $H^{2'}$), 5.24 (t, *J* = 5.3 Hz, 1H, $H^{3'}$), 4.32 (q, *J* = 4.2 Hz, 3H, $H^{4'}$, $H^{5'}$), 2.40 – 2.17 (m, 6H, -COC-CH_2CH$_2$-), 1.71 – 1.05 (m, 54H, -CH_2-), 0.91 – 0.70 (m, 9H, -CH_3).



C2.1. A suspension of **C2.2** (2.79 mmol, 2.20 g), I_2 (2.5 mmol, 0.6 g) and HIO_3 (4.7 mmol, 0.8 g) in 68 mL acetic acid was stirred overnight at 40 ^{o}C. Once the reaction was completed, the insoluble HIO_3 was filtered and discarded. A 1:1 AcOEt/Et_2O mixture (25 mL) was added to the reaction mixture and this was washed with 3 x 50 mL water, 3 x 50 mL $NaHCO_3$ (sat), 1 x 50 mL $Na_2S_2O_3$ (sat) and finally 1 x 50 mL water. The organic layer was dried over $MgSO_4$ and the solvent was evaporated. The product was then purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (40:1), obtaining **C2.1** as a yellow oil (1.86 g, 63%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.90 (s, 1H, H^6), 6.08 (d, *J* = 3.9 Hz, 1H, $H^{1'}$), 5.32 (t, *J* = 4.3 Hz, 1H, $H^{2'}$), 4.46 – 4.27 (m, 4H, $H^{3'}$,$H^{4'}$, $H^{5'}$), 2.33 (q, *J* = 7.2 Hz, 6H, -COC-CH_2CH$_2$-), 1.79 – 1.10 (m, 54H, -CH_2-), 0.94 – 0.81 (m, 9H, -CH_3).



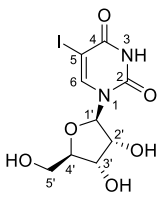
C2. **C2** was prepared following the *Standard Procedure A* for a Sonogashira coupling. **C2.1** (2.1 mmol, 2 g), $PdCl_2(PPh_3)_2$ (0.044 mmol, 30 mg) and CuI (0.02 mmol, 4.2 mg) were dissolved in the THF/Et_3N (4:1) mixture (30 mL). Then TMSA (3.27 mmol, 0.32 g) was added and the mixture was stirred at 40 ^{o}C during 24 h. The product was used in the next reaction without purification. Following *Standard Procedure B*, TBAF·3H_2O (3.2 mmol, 1.0 g) was added over a THF (20 mL) solution of the previous crude product. **C2** was purified by chromatography on silica gel eluted by $CHCl_3$/MeOH; (40:1). The product was then purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (40:1), obtaining **C2** as a yellow solid (0.61 g, 80%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.95 (s, 1H, H^6), 6.11 (s, 1H, $H^{1'}$), 5.77 (s (broad), 1H, NH_2), 5.34 (d, *J* = 14.9 Hz, 2H, $H^{2'}$, $H^{3'}$), 4.37 (s, 3H, $H^{4'}$,$H^{5'}$), 3.35 (s, 1H, -CC*H*), 2.36 (ddt, *J* = 29.2, *J'* = 15.3, *J"* = 7.5 Hz, 6H, -COC-CH_2CH$_2$-), 1.28 (d, *J* = 10.0 Hz, 54H, -CH_2-), 0.88 (t, *J* = 6.5 Hz, 9H, -CH_3).

^{13}C-NMR (75 MHz, DMSO-d_6) δ(ppm) = 173.3, 172.5, 172.1, 163.2, 153.8, 143.2, 131.3, 129.3, 124.6, 94.4, 91.4, 88.9, 82.3, 79.5, 73.9, 69.2, 62.4, 34.2, 33.8, 31.6, 30.4, 29.8, 29.7, 29.6, 29.6, 29.6, 29.4, 29.3, 29.3, 29.1, 29.1, 24.9, 24.7, 24.5, 22.3, 14.8.

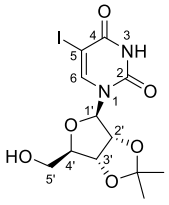
HRMS (MALDI): Calculated for $C_{47}H_{79}N_3O_8$: 814.1620 $[M+H]^+$. Found: 836.5763 $[M+Na]^+$. Matrix: DCTB

Synthesis of the uridine derivatives.



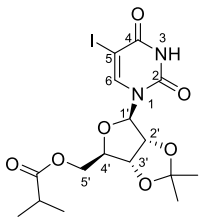
U1.3. U1.3 was synthesized according to a literature procedure.[139b] The pristine uridine (40.95 mmol, 10 g) was dissolved in MeCN (250 mL) and stirred. Then, iodine (24.57 mmol, 6.23 gr) and CAN (20.47 mmol, 11.22 gr) were added and stirred at 40 ^{o}C during 5 hours. After reaction completion, the solvent was removed and the crude was extracted whith CHCl3 (3 x 100ml) and NaCl (2 x 100 mL). The organic phase was dried with $MgSO_4$, filtered, and concentrated. **U1.3** was obtained as a pale yellow solid. (12.5 g, 84%).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.65 (s (broad), 1H, N*H*), 8.21 (s, 1H, H^6); 5.79 (d, *J* = 2.1 Hz, 1H, $H^{1'}$), 5.21 (t, *J* = 5.2 Hz, 1H, CH_2-O*H*), 4.85 (dd, *J* = 6.5 Hz, *J'* = 2.6 Hz, 1H, $H^{2'}$), 4.69 (dd, *J* = 6.5 Hz, *J'* = 3.5 Hz, 1H, $H^{3'}$), 4.12 – 4.03 (m, 1H, $H^{4'}$), 3.71 – 3.59 (m, 2H, $H^{5'}$).



U1.2. **U1.2** was synthesized according to a literature procedure[135b] adapted to our molecule. Iodouridine **U1.3** (67.55 mmol, 25.0 g) was made to react with $HClO_4$ (70%) (1.88 eq, 101.32 mmol, 8.76 mL) in acetone (1000 mL). After stirring for 30 min, the reaction was completed. Then, dry $CaCO_3$ was added and the resulting mixture was stirred overnight. The reaction mixture was then filtered through a short silica plug (acetone/$CHCl_3$; (3:1)) and the product was collected after evaporation of the solvent. The product was purified by recrystallization in AcOEt to give a white solid. (24,06 g, 87%).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.73 (s (broad), 1H, N*H*), 8.33 (s, 1H, H^6); 5.82 (d, *J* = 2.1 Hz, 1H, $H^{1'}$), 5.17 (t, *J* = 5.2 Hz, 1H, CH_2-O*H*), 4.92 (dd, *J* = 6.5 Hz, *J'* = 2.6 Hz, 1H, $H^{2'}$), 4.75 (dd, *J* = 6.5 Hz, *J'* = 3.5 Hz, 1H, $H^{3'}$), 4.15 – 4.05 (m, 1H, $H^{4'}$), 3.68 – 3.52 (m, 2H, $H^{5'}$), 1.48 (s, 3H, -OC(CH_3)), 1.29 (s, 3H, -OC(CH_3)).

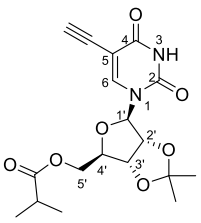


U1.1. To a solution of **U1.2** (12.19 mmol, 5.0 g) and catalitic DMAP (0.2 eq, 2.44 mmol, 298 mg) in dry MeCN (120 mL), NEt_3 (1.5 eq, 18.29 mmol, 2.54 mL) and isobutyric anhydride (1.1 eq, 13.41 mmol, 2.23 mL) were added. The reaction mixture was stirred at room temperature overnight. MeOH (2 mL) was then added the mixture further stirred for 30 min. The solvent was evaporated under reduced pressure. Then, the mixture was dissolved in $CHCl_3$ and washed with $NaHCO_3$ (sat) (2 x 100 mL) and NaCl (2 x 100 mL). The organic phase was dried with $MgSO_4$, filtered, and concentrated. The product was purified by column chromatography in $CHCl_3$/MeOH (60:1). 5.04 g of compound **U1.1** was obtained (86% yield).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.81 (s (broad), 1H, N*H*), 8.14 (s, 1H, H^6), 5.78 (s, 1H, $H^{1'}$), 5.06 (dd, J = 6.4 J' = 1.8 Hz , 1H, $H^{2'}$), 4.78 (d, J = 6.2 Hz, 1H, $H^{3'}$), 4.35 – 4.10 (m, 3H, $H^{4'}$, $H^{5'}$), 2.55 – 2.49 (m, 1H, -COC*H*), 1.48 (s, 3H, -OC(CH_3)), 1.29 (s, 3H, -OC(CH_3)), 1.11 (s, 3H, -COCH-(CH_3)$_2$), 1.09 (s, 3H, -COCH-(CH_3)$_2$).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 176.5, 160.2, 149.8, 146.1, 114.8, 94.1, 85.1, 84.7, 80.8, 68.7, 63.8, 33.9, 27.1, 25.3, 19.1, 19.0.

MS (FAB+): 481.1 $[M+H]^+$.

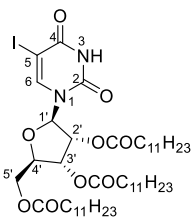


U1. Following *Standard Procedure A*, to a solution of compound **U1.1** (16.65 mmol, 8 g), $PdCl_2(PPh_3)_2$ (0.33 mmol, 0.234 g) and CuI (0.17 mmol, 32 mg) in NEt_3/THF (4:1) (80 mL), TMSA (1.5 eq, 24.98 mmol, 3.83 mL) was added and the mixture was stirred at 40 ^{o}C for 18 h. After reaction completion, the deprotection reaction is carried out without further purification following the *Standard Procedure* A. The crude product obtained after solvent evaporation was suspended in dry THF (100 mL) and TBAF·3H_2O (1.2 eq, 19.98 mmol, 6.30 g) was added. After reaction completion the solvent was evaporated and the resulting residue was purified by column chromatography using $CHCl_3$/ AcOEt (5:1) as eluent. 4.85 g of compound **U1** was obtained as a white solid (77% yield).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.76 (s (broad), 1H, N*H*), 8.05 (s, 1H, H^6), 5.81 (d, J = 1.9 Hz, 1H, $H^{1'}$), 5.06 (dd, J = 6.5, J' = 1.9 Hz, 1H, $H^{2'}$), 4.79 (d, J = 6.4 Hz, 1H, $H^{3'}$), 4.33 – 4.17 (m, 3H, $H^{4'}$, $H^{5'}$), 4.14 (s, 1H, -CC*H*), 2.64 – 2.53 (m, 1H, -COC*H*), 1.49 (s, 3H, -OC(CH_3)), 1.29 (s, 3H, -OC(CH_3)), 1.09 (dd, J = 7.0, J' = 1.7 Hz, 6H, -COCH-(CH_3)$_2$).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 176.5, 161.2, 148.9, 144.7, 114.8, 99.4, 93.9, 85.1, 84.8, 82.4, 80.6, 74.1, 63.7, 33.9, 27.1, 25.3, 19.1, 18.9.

MS (FAB+): 379.1 $[M+H]^+$.

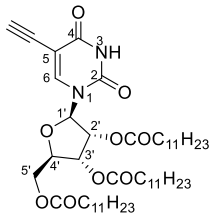


U2.1. The iodinated **U1.3** (33.80 mmol, 12.5 g) and a catalytic amount of DMAP (0.6 eq, 20.28 mmol, 2.48 g) were dissolved in 300 mL of dry THF/DMF (2:1). Then NEt_3 (4.5 eq, 152.1 mmol, 15.40 mL) and lauroyl chloride (3.1 eq, 104.78 mmol, 24.23 mL) were added. The mixture is stirred at room temperature overnight. The reaction is followed by TLC and, once finished, 10 mL MeOH were added and the mixture is further stirred for 30 min in order to react with the remaining chloride. The solvent was removed and the solid was extracted with water (3 x 100 mL). The organic layer was dried over $MgSO_4$ and the solvent is evaporated. The product was then purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), obtaining a yellow oil (25.1 g, 81%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.32 (s (broad), 1H, N*H*), 7.90 (s, 1H, H^6), 6.09 (d, J = 5.3 Hz, 1H, $H^{1'}$), 5.30 (q, J = 5.7, J' = 5.0 Hz, 2H, $H^{2'}$, $H^{3'}$), 4.36 (dt, J = 23.4, J' = 12.4 Hz, 3H, $H^{4'}$, $H^{5'}$), 3.66 (s, 1H, -CC*H*), 2.58 – 2.21 (m, 6H, -OCH_2CH$_2$-), 1.76 – 1.15 (m, 54H, -CH_2-), 0.98 – 0.77 (m, 9H, -CH_3).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 174.3, 172.9, 172.4, 159.2, 149.6, 143.7, 86.9, 80.7, 73.0, 70.2, 69.5, 62.9, 51.4, 34.2, 34.1, 33.9, 33.7, 31.9, 29.6, 29.6, 29.5, 29.3, 29.3, 29.1, 29.1, 29.1, 25.0, 24.8, 24.6, 22.7, 14.1.

HRMS (MALDI): Calculated for $C_{45}H_{77}IN_2O_9$: 917.0205 $[M+H]^+$. Found: 939.4556 $[M+Na]^+$. Matrix: DCTB



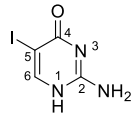
U2. Following *Standard Procedure A*, to a solution of compound **U2.1** (16.65 mmol, 16 g), $PdCl_2(PPh_3)_2$ (0.33 mmol, 0.234 g) and CuI (0.17 mmol, 32 mg) in THF/NEt_3 (4:1) (80 mL) TMSA (1.5 eq, 24.98 mmol, 3.83 mL) was added and the mixture was stirred at 40 ^oC for 18 h. After reaction completion, the deprotection reaction is carried out without further purification following the *Standard Procedure* A. The crude product obtained after solvent evaporation was suspended in dry THF (100 mL) and TBAF·$3H_2O$ (1.3 eq, 21.6 mmol, 6.81 g) was added. After reaction completion the solvent was evaporated and the resulting residue was purified by column chromatography using $CHCl_3$/MeOH (60:1) as eluent. 9.36 g of compound **U2** was obtained as a white solid (71% yield).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 9.46 (s (broad), 1H, N*H*), 7.82 (s, 1H, H^6), 6.08 (d, J = 4.3 Hz, 1H, $H^{1'}$), 5.30 (d, J = 3.4 Hz, 2H, $H^{2'}$, $H^{3'}$), 4.48 – 4.23 (m, 3H, $H^{4'}$, $H^{5'}$), 3.15 (s, 1H, -CC*H*), 2.53 – 2.24 (m, 6H, -OCH_2CH_2-), 1.73 – 1.12 (m, 54H, -CH_2-), 0.84 (t, J = 6.5 Hz, 9H, -CH_3).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 172.9, 172.4, 172.3, 160.7, 149.0, 142.6, 100.2, 87.2, 82.4, 80.6, 74.3, 73.1, 70.0, 62.8, 34.1, 33.9, 33.7, 31.9, 29.6, 29.5, 29.3, 29.3, 29.1, 29.1, 24.8, 24.7, 22.7.

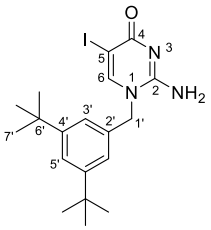
HRMS (MALDI): Calculated for $C_{47}H_{78}N_2O_9$: 815.1460 $[M+H]^+$. Found: 837.5588 $[M+Na]^+$. Matrix: DCTB

Synthesis of the *iso*cytidine derivatives.



iC1.2. The synthesis of **iC1.2** was performed according to a published procedure.[139c] The pristine *iso*cytidine (82.8 mmol, 9.20 gr) was dissolved in acetic acid/H_2O (1:1) and the mixture was stirred at 50 ^oC. Then, NIS was added and the reaction was heated at 100 ^oC for 40 min. After reaction completion, the crude was filtered and washed with H_2O affording **iC1.2** as a white solid (15.6 gr, 84%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 11.21 (s (broad), 1H, N*H*), 8.00 (s, 1H, H^6), 6.71 (s (broad), 1H, NH_2).



iC1.1. The synthesis of **iC1.1** was performed according to a published procedure[138] that was adapted to our molecule. To a solution of **iC1.2** (13.15 mmol, 3.05 g) in DMF (50 mL) 3,5-di-*tert*-butylbenzyl bromide (14.47 mmol, 4.09 g) and tetrabuthylamonium hydroxide (15.78 mmol, 4.09 g) were added. The mixture was heated to 40 ^oC. After 2 hours, the solvent was evaporated under vacuum and the mixture was redissolved in $CHCl_3$ and washed with water (2 x 50 mL). The organic phase was dried over $MgSO_4$, the solvent evaporated and the residue purified by column chromatography using $CHCl_3$/MeOH (30:1) as eluent. 4.16 g of compound **iC2** was obtained (72% yield) as a yellow oil.

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.48 (s, 1H, H^6), 7.17 (s, 1H, $H^{5'}$), 7.10 (s (broad), 2H, NH_2), 6.84 (d, 2H, J = 1.8 Hz, $H^{3'}$), 4.82 (s, 2H, $H^{1'}$), 1.07 (s, 18H, -C(CH_3)$_3$).

^{13}C-NMR (75.0 MHz, DMSO-d_6) δ(ppm) = 167.4, 163.3, 162.0, 157.2, 155.8, 151.8, 147.4, 133.1, 122.5, 121.2, 78.0, 69.3, 54.4, 34.9, 31.4.

MS (FAB+): 440.1 $[M+H]^+$.

Since *iso*cytidine has the capability to form tautomeric species in solution, it will lead the reaction to form three different derivatives. Once finished the reaction, as each one of the compounds reveal different polarities and ^{1}H NMR spectra (see Figure 1.6), we could purified and characterized them easily. Figure 1.6b,c shows different alkylated derivatives and its respective ^{1}H NMR spectra in which we can observe directly how much is affected the benzylic CH_2 shifts. Because of that, the O-alkylated molecule possess the most down-shifted CH_2 signal and the lowest polarity of all of them. In contrast, as it is shown in Figure1.6, the N(1)-substituted targeted molecule displays the most up-shifted signal for the benzylic CH_2 and possess the most polarity among all derivatives. Finally, Figure 1.6d shows clearly that our suggestions were true due to presence of strong cross-peaks in a NOESY spectra between the benzylic CH_2 and the 6 position in the pyrimidine ring.

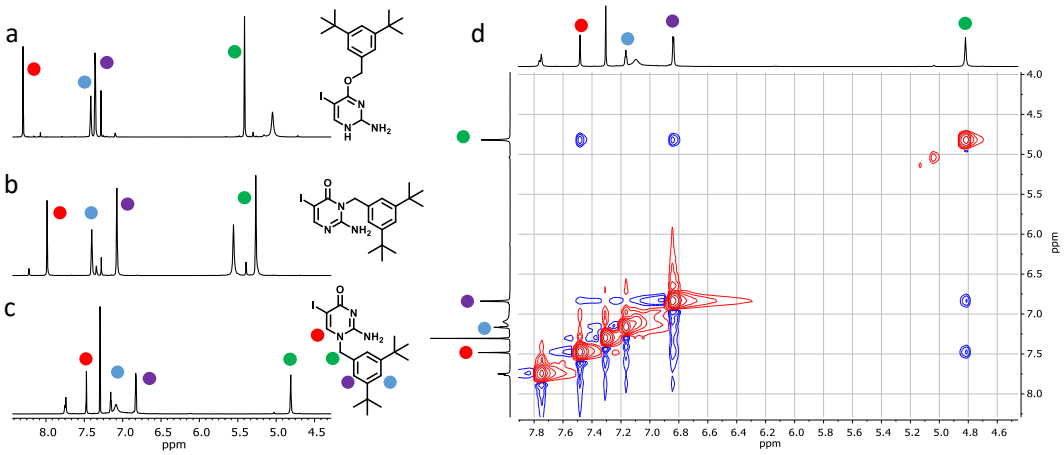
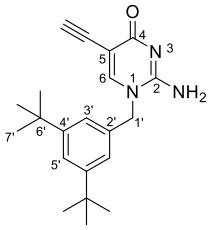


Figure 1.6. ^{1}H NMR spectra of the different compounds obtained through the alkylation step. (**a**) O-substituted, (**b**) N(3)-substituted, (**c**) N(1)-substituted, and (**d**) the NOESY ^{1}H-NMR spectra of the desired compound.

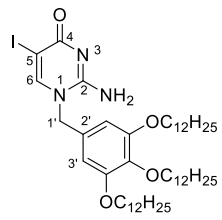


iC1. Following *Standard Procedure A*, over a solution of **iC2** (6.12 mmol, 2.69 g), $PdCl_2(PPh_3)_2$ (0.122 mmol, 85.9 mg) and CuI (0.061 mmol, 11.66 mg) in THF/NEt_3 (4:1) (120 mL) were added. Then, TMSA (9.18 mmol, 0.902 g) was added. The crude was directly deprotected with TBAF $3H_2O$ (7.34 mmol, 2.31 g) using *Standard Procedure A*. The compound was purified by column chromatography in $CHCl_3$/MeOH (30:1). Recrystallization in MeCN finally yielded **iC1** (1.04 g, 68%) as a pale solid.

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.27 (s, 1H, H^6), 7.22 (s, 1H, $H^{5'}$), 6.90 (d, 2H, J = 1.7 Hz, $H^{3'}$), 6.49 (s (broad), 2H, NH_2), 4.82 (s, 2H, $H^{1'}$), 2.97 (s, 1H, -CCH), 1.15 (s, 18H, -C(CH_3)$_3$).

^{13}C-NMR (75 MHz, DMSO-d_6) δ(ppm) = 150.7, 147.3, 134.6, 121.4, 121.1, 82.6, 78.5, 53.1, 40.4, 40.1, 39.8, 39.5, 39.2, 39.0, 38.7, 34.51, 31.2, 23.1, 19.2, 13.5.

MS (FAB+): 338.2 $[M+H]^+$.

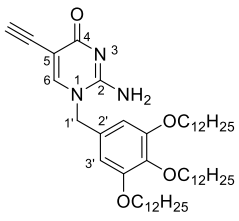


iC2.1. The synthesis of **iC2.1** was performed according to a published procedure[138] that was adapted to our molecule. To a solution of **iC1.2** (4.34 mmol, 1 g) in DMF (10 mL), 3,5-di-*tert*-butylbenzyl bromide (4.83 mmol, 3.5 g) and tetrabuthylamonium hydroxide (5.2 mmol, 5.2 ml (1M)) were added. The mixture was heated to 40 °C. After 2 hours, the solvent was evaporated under vacuum and the mixture was redissolved in $CHCl_3$ and washed with water (2 x 50 mL). The organic phase was dried over $MgSO_4$, the solvent evaporated and the residue purified by column chromatography using $CHCl_3$/MeOH (150:1) as eluent. 2.67 g of compound **iC2** was obtained (70% yield) as an orange oil.

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.02 (s, 1H, H^6), 6.42 (s (broad), 2H, NH_2), 5.16 (s, 2H, $H^{3'}$), 5.06 (s, 2H, $H^{1'}$), 3.92 (t, J = 6.5 Hz, 6H, -OCH_2CH$_2$-), 1.74 – 1.16 (m, 60H, -CH_2-), 0.98 – 0.79 (m, 12H, -CH_3).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 152.1, 146.5, 142.7, 136.5, 129.4, 118.0, 105.2, 98.9, 71.9, 67.8, 30.7, 29.2, 28.5, 28.3, 28.2, 25.0, 21.5, 13.2.

MS (MALDI): 902.6 $[M+Na]^+$. Matrix: DCTB

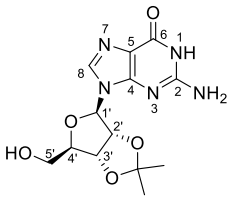


iC2. Following *Standard Procedure A*, over a solution of **iC2.1** (5.68 mmol, 0.5 g), $PdCl_2(PPh_3)_2$ (11.4 µmol, 8 mg) and CuI (5.7 µmol, 1 mg) in THF/NEt_3 (4:1) (10 mL) were added. Then, TMSA (1.13 mmol, 0.17 mL) was added. The crude was directly deprotected with TBAF $3H_2O$ (0.85 mmol, 0.27 g) using *Standard Procedure A*. The compound was purified by column chromatography in $CHCl_3$/MeOH (100:1). Recrystallization in MeCN finally yielded **iC2** (0.36 g, 82%) as a pale solid.

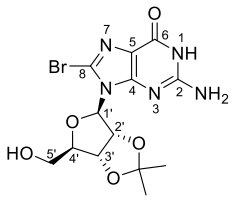
^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.09 (s, 1H, H^6), 6.25 (s, 2H, $H^{3'}$), 5.01 (s, 2H, $H^{1'}$), 4.70 (s (broad), 2H, NH_2), 3.73 (d, J = 6.8 Hz, 6H, -OCH_2CH$_2$-), 2.95 (s, 1H, -CCH), 1.55 (p, J = 7.3 Hz, 6H, -OCH_2CH$_2$-), 1.10 (d, J = 10.1 Hz, 60H, -CH_2-), 0.71 (d, J = 6.3 Hz, 9H, -CH_3).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 153.9, 146.2, 138.6, 106.2, 73.5, 69.3, 55.2, 31.9, 30.4, 29.8, 29.7, 29.7, 29.5, 29.5, 29.4, 26.2, 26.2, 22.7, 14.1.
HRMS (ESI+): Calculated for $C_{49}H_{83}N_3O_4$: 779.2200 $[M+H]^+$. Found: 800.6299 $[M+Na]^+$.

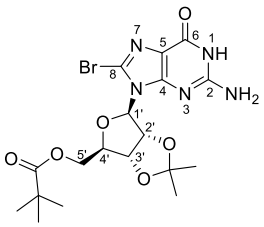
Synthesis of the guanosine derivatives.



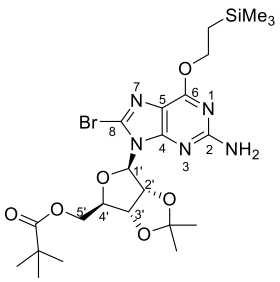
G1.4. The synthesis of **G1.4** was performed according to a published procedure.[135] The pristine guanosine (88.26 mmol, 25.0 g) was made to react with $HClO_4$ (70%) (1.88 eq, 166 mmol, 10.02 mL) in acetone (1000 mL). After stirring for 30 min, the reaction was completed. Then, $NH_{3\ (aq)}$ (10.02 mL) was added until the resulting mixture precipitated. The reaction mixture was then filtered through a short silica plug (acetone/$CHCl_3$; (3:1)) and the product was collected after evaporation of the solvent. Afterwards, the crude was filtered and washed with cold H_2O. **G1.4** was obtained as a white solid (24.12 gr, 93%).
^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = δ 10.67 (s (broad), 1H, N*H*),7.92 (s, 1H, NC-*H*), 6.50 (s (broad), 2H, NH_2), 5.93 (d, *J* = 2.7 Hz, 1H, $H^{1'}$), 5.19 (t, *J* = 4.1 Hz, 1H, $H^{2'}$), 5.11 – 4.80 (m, 2H, $H^{3'}$), 4.12 (s, 2H, $H^{4'}$, $H^{5'}$), 1.52 (s, 3H, -OC(CH_3)), 1.32 (s, 3H, -OC(CH_3)).



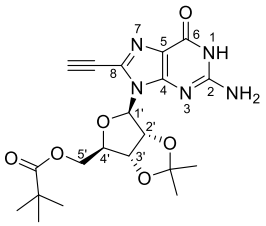
G1.3. The synthesis of **G1.3** was performed according to a published procedure.[134b] **G1.4** (88.26 mmol, 28.52 g) was suspended in H_2O/MeCN (1:3) and NBS (88.26 mmol, 15.7 g) was added in portions during 1 hour and stirred at room temperature. After 2 hours, the crude was filtered and washed with cold acetone to afford **G1.3** as a pale yellow solid (31.15 gr, 89%).
^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.91 (s (broad), 1H, N*H*), 6.72 (s (broad), 2H, NH_2), 5.89 (d, *J* = 1.2 Hz, 1H, $H^{1'}$), 5.43 (d, *J* = 6.0 Hz, 1H, $H^{2'}$), 5.23 (dd, *J* = 3.6, *J'* = 6.0 Hz, 1H, $H^{3'}$), 4.28 – 4.12(m, 2H, $H^{5'}$), 4.08 – 3.93 (m, 1H, $H^{4'}$), 1.50 (s, 3H, -OC(CH_3)), 1.33 (s, 3H, -OC(CH_3)).



G1.2. In a 500-mL round-bottomed flask, equipped with a magnetic stirrer, **G1.3**[135b] (12.0 mmol, 4.83 g) and DMAP (2.4 mmol, 290 mg) were placed. Dry DMF (400 mL) was added and the mixture was stirred at room temperature under argon until the solid was dissolved. Then, NEt_3 (18 mmol, 2.5 mL) and trimethylacetic anhydride (36.0 mmol, 7.30 mL) were added. The resulting mixture was stirred at 130 ^{o}C until **G1.3** was consumed. Afterwards, MeOH (7 mL) was added and the mixture was stirred during 15 minutes. The solvent was then eliminated under reduced pressure and the solid was directly purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (30:1) and then by recrystallization (CH_2Cl_2/hexane). We obtained a white solid (4.32 g, 63%).
^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.89 (s (broad), 1H, N*H*), 6.68 (s (broad), 2H, NH_2), 5.91 (d, *J* = 1.3 Hz, 1H, $H^{1'}$), 5.45 (d, *J* = 6.2 Hz, 1H, $H^{2'}$), 5.27 (dd, *J* = 3.8, *J'* = 6.0 Hz, 1H, $H^{3'}$), 4.22 – 4.10 (m, 2H, $H^{5'}$), 4.01 – 3.89 (m, 1H, $H^{4'}$), 1.51 (s, 3H, -OC(CH_3)), 1.32 (s, 3H, -OC(CH_3)), 1.09 (s, 9H, -COC-(CH_3)$_3$).
^{13}C-NMR, (75 MHz, DMSO-d_6) δ(ppm) = 177.1, 155.6, 153.8, 151.4, 120.2, 117.0, 113.2, 89.8, 85.6, 83.2, 81.4, 64.0, 40.4, 26.9, 26.8, 25.3.
HRMS (ESI+): Calculated for $C_{18}H_{25}BrN_5O_6$: 486.0910 $[M+H]^+$. Found: 486.0979 $[M+H]^+$.



G1.1. In a 50-mL round-bottomed flask, equipped with a magnetic stirrer, **G1.2** (1.70 mmol, 1.17 g), PPh_3 (2.55 mmol, 668.8 mg) and DIAD (2.38 mmol, 0.47 mL) were placed. Dry dioxane (15 mL) was added and the mixture was stirred at room temperature under argon atmosphere until the solid was dissolved. Then 2-trimethylsilylethanol was added dropwise (2.72 mmol, 0.39 mL) and the mixture was stirred at room temperature during 12 h. Finally, the solvent was eliminated under reduced pressure and the oil obtained was purified by chromatography on silica gel eluted with Hexane/AcOEt (6:1) to yield a brown solid (1.50 g, 99%).
^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.70 (s (broad), 2H, NH_2), 5.91 (d, *J* = 1.2 Hz, 1H, $H^{1'}$), 5.45 (d, *J* = 6.0 Hz, 1H, $H^{2'}$), 5.27 (dd, *J* = 3.6, *J'* = 6.0 Hz, 1H, $H^{3'}$), 4.50 (t, *J* = 8.2 Hz, 2H, CO-CH_2-), 4.31 – 4.19 (m, 2H, $H^{5'}$), 4.11 – 4.00 (m, 1H, $H^{4'}$), 1.51 (s, 3H, -OC(CH_3)), 1.32 (s, 3H, -OC(CH_3)), 1.09 (s, 9H, -COC-(CH_3)$_3$), 0.06 (s, 9H, -Si(CH_3)$_3$).

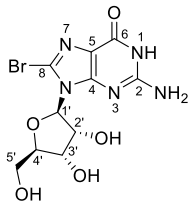


G1. G1 was obtained following *Standard Procedure A* using **G1.1** (1.65 mmol, 0.97 g), $PdCl_2(PPh_3)_2$ (0.03 mmol, 23.0 mg), CuI (0.01 mmol, 4.0 mg), TMSA (5 mmol, 0.62 g) and THF/ NEt_3 (20 mL) were mixed. The mixture was then stirred at 40 ºC during 24 h. The crude was directly deprotected with TBAF·$3H_2O$ (4 mmol, 1.26 g) and THF (20 mL). After solvent evaporation the brown oil was purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1). **G1** was obtained as a pale solid (0.563 g, 44%).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.02 (s (broad), 1H, N*H*), 6.77 (s (broad), 2H, NH_2), 6.02 (d, *J* = 1.3 Hz, 1H, $H^{1'}$), 5.45 – 5.32 (d, *J* = 6.4 Hz, 1H, $H^{2'}$), 5.22 (m, 1H, $H^{3'}$), 4.84 (s, 1H, -C*H*), 4.26-4.10 (m, 3H, $H^{4'}$, $H^{5'}$), 1.51 (s, 3H, -OC(CH_3)), 1.31 (s, 3H, -OC(CH_3)), 1.09 (s, 9H, -COC-(CH_3)$_3$).

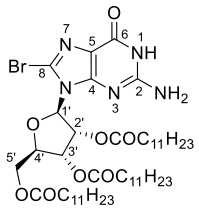
^{13}C-NMR, (75 MHz, DMSO-d_6) δ(ppm) = 177.2, 156.2, 154.3, 150.1, 128.4, 116.7, 113.3, 88.6, 86.0, 85.4, 83.4, 81.4, 72.9, 64.2, 40.4, 40.1, 40.0, 39.8, 39.5, 39.2, 39.0, 38.7, 38.2, 27.0, 26.8, 25.3.

HRMS (ESI+): Calculated for $C_{20}H_{26}N_5O_6$: 432.1805 $[M+H]^+$. Found: 432.1889 $[M+H]^+$.



G2.3. The synthesis of **G2.3** was performed according to a published procedure.[134b] The pristine guanosine (88.26 mmol, 25 g) was suspended in H_2O/MeCN (1:3) and NBS (88.26 mmol, 15.7 g) was added in portions during 1 hour and stirred at room temperature. After 2 hours, the crude was filtered and washed with cold acetone to afford **G2.3** as a pale yellow solid (25.8 gr, 81%).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.93 (s (broad), 1H, N*H*), 6.75 (s (broad), 2H, NH_2), 5.83 (d, *J* = 1.0 Hz, 1H, $H^{1'}$), 5.48 (d, *J* = 6.1 Hz, 1H, $H^{2'}$), 5.20 (dd, *J* = 3.5, *J'* = 6.1 Hz, 1H, $H^{3'}$), 4.22 – 4.10(m, 2H, $H^{5'}$), 4.06 – 3.91 (m, 1H, $H^{4'}$).

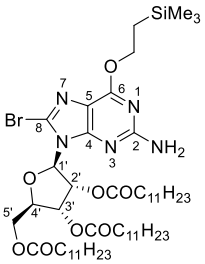


G2.2. In a 500-mL round-bottomed flask, equipped with a magnetic stirrer, **G2.3** (11.0 mmol, 4 g) and DMAP (2.2 mmol, 267 mg) were placed. Dry THF/DMF (2:1) (40 mL) was added and the mixture was stirred at room temperature under argon until the solid was dissolved. Then, NEt_3 (17 mmol, 2.4 mL) and lauroyl chloride (33.0 mmol, 7.62 mL) were added. The resulting mixture was stirred at 130 ºC until **G2.3** was consumed. Afterwards, MeOH (7 mL) was added and the mixture was stirred during 15 minutes. The solvent was then eliminated under reduced pressure and the solid was directly purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (60:1). **G2.2** was obtained as a white solid (6.80 g, 68%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 11.94 (s (broad), 1H, N*H*), 6.30 (s (broad), 2H, NH_2), 5.94 (s, 2H, $H^{1'}$, $H^{2'}$), 4.41 (d, *J* = 44.7 Hz, 4H, $H^{3'}$, $H^{4'}$, $H^{5'}$), 2.58 – 2.04 (m, 6H, -OCOCH_2-), 1.77 – 1.09 (m, 54H, -CH_2-), 0.86 (t, 9H, -CH_3).

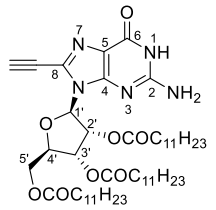
^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 179.2, 173.6, 172.3, 172.1, 172.0, 158.4, 157.7, 153.4, 152.4, 121.9, 117.8, 116.7, 88.4, 79.7, 72.0, 70.3, 62.9, 34.1, 34.0, 33.9, 33.8, 31.9, 29.6, 29.4, 29.3, 29.3, 29.1, 25.1, 24.8, 22.7, 14.1.

HRMS (MALDI): Calculated for $C_{46}H_{78}BrN_5O_8$: 910.0612 $[M+H]^+$. Found: 932.4917 $[M+Na]^+$. Matrix: DCTB



G2.1. In a 250-mL round-bottomed flask, equipped with a magnetic stirrer, **G2.2** (7.48 mmol, 6.8 g), PPh_3 (28.61 mmol, 7.5 g) and DIAD (6.27 mmol, 1.24 mL) were placed. Dry dioxane (140 mL) was added and the mixture was stirred at room temperature under argon atmosphere until the solid was dissolved. Then 2-trimethylsilylethanol was added dropwise (11.96 mmol, 1.72 mL) and the mixture was stirred at room temperature during 12 h. Finally, the solvent was eliminated under reduced pressure and the oil obtained was purified by chromatography on silica gel eluted with Hexane/AcOEt (10:1) to yield a brown solid (5.97 g, 79%)

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 6.19 (s, 1H, $H^{1'}$), 5.01 (s (broad), 2H, NH_2), 4.66 – 4.33 (m, 7H, $H^{2'}$, $H^{3'}$, $H^{4'}$, $H^{5'}$, CO-CH_2-,), 2.43 – 2.24 (m, 6H, -OCOCH_2-), 1.82 – 0.79 (m, 56H, -CH_2-, -CH_2-Si-(CH_3)$_3$), 0.32 (s, 9H, -COC-(CH_3)$_3$), 0.12 (s, 9H, -Si(CH_3)$_3$).



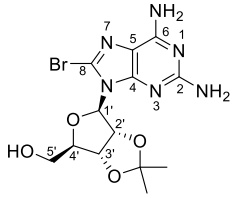
G2. **G2** was obtained following *Standard Procedure A* using **G2.1** (5.9 mmol, 5.97 g), $PdCl_2(PPh_3)_2$ (0.06 mmol, 46.0 mg), CuI (0.02 mmol, 8.0 mg), TMSA (10 mmol, 1.2 g) and THF/ NEt_3 (4:1, 80 mL) were mixed. The mixture was then stirred at 40 ^oC during 24 h. The crude was directly deprotected with TBAF $3H_2O$ (8 mmol, 2.52 g) and THF (80 mL). After solvent evaporation the brown oil was purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1). **G2** was obtained as a brown solid (3.07 g 61%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 10.98 (s (broad), 1H, N*H*), 7.52 (s (broad), 2H, NH_2), 6.07 (s, 2H, $H^{1'}$), 5.96 (d, J = 5.0 Hz, 1H, $H^{2'}$), 5.72 (s, 1H, $H^{3'}$), 4.46 (d, J = 10.8 Hz, 1H, $H^{4'}$), 4.38 – 4.14 (m, 2H, $H^{5'}$), 3.8 (s, 1H, -CC*H*), 2.23 (dt, J = 22.2, J' = 7.7 Hz, 6H, -OCOCH_2-), 1.70 – 0.96 (m, 54H, -CH_2-), 0.75 (t, 9H, -CH_3).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 179.8, 173.4, 172.3, 172.1, 172.0, 158.4, 157.7, 153.4, 152.4, 121.9, 117.8, 116.7, 88.4, 79.7, 72.0, 70.3, 62.9, 34.3, 34.0, 31.9, 29.5, 29.2, 24.8, 22.7, 14.1.

HRMS (MALDI): Calculated for $C_{48}H_{79}N_5O_8$: 854.6007 $[M+H]^+$. Found: 876.5821 $[M+Na]^+$. Matrix: DCTB

Synthesis of the 2-aminoadenosine derivatives

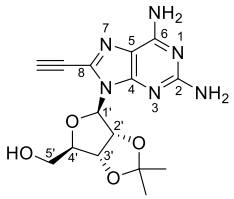


A1.2. **A1.3**[135a] (62.1 mmol, 20.0 g) was dissolved in a MeCN/H_2O (4:1) (500 mL) solvent mixture and N-bromosuccinimide (NBS; 84.3 mmol, 15.0 g) was added in three portions, the mixture was stirred for 2 h at room temperature. Once the reaction was completed, MeCN was removed under vacuum pressure, and $NaHCO_3$ (sat) was added until a yellow precipitate appeared. This solid was then filtered, reprecipitated with cold MeCN and purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1). Product **A1.2** was obtained as a white solid (12.4 g, 50%).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.95 (s, 2H, NH_2), 6.00 (s, 2H, NH_2), 5.89 (d, J = 2.1 Hz, 1H, $H^{1'}$), 5.51 (dd, J = 6.2, J' = 2.1 Hz, 1H, $H^{2'}$), 5.13 (dd, J = 6.3, J' = 3.3 Hz, 1H, $H^{3'}$), 5.02 (t, J = 5.8 Hz, 1H, O*H*), 4.10 (m, J = 6.1, 1H, H$^{4'}$), 3.61 – 3.40 (m, 2H, -CH_2-OH), 1.53 (s, 3H, -OC(CH_3)), 1.33 (s, 3H, -OC(CH_3)).

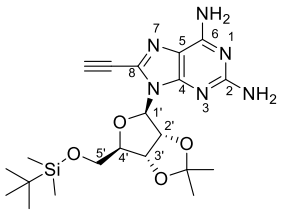
^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 159.2, 155.4, 151.3, 122.6, 115.1, 114.0, 93.4, 85.6, 82.2, 81.7, 63.4, 27.8, 25.6.

HRMS (ESI+): Calculated for $C_{13}H_{18}BrN_6O_4$: 401.0495 $[M+H]^+$. Found: 401.0582 $[M+H]^+$



A1.1. A1.1 was prepared following *Standard Procedure A*. **A1.2** (21.0 mmol, 8.42 g), $PdCl_2(PPh_3)_2$ (0.4 mmol, 281 mg), CuI (0.2 mmol, 38.0 mg), TMSA (30.0 mmol, 4.9 g) were mixed in the THF/NEt_3 solvent (20 mL). The mixture was then stirred at 40 ^oC during 24 h. After removal of the solvent, a brown oil was obtained that was used in the following reaction step without previous purification. Then, following *Standard Procedure A*, the brown oil was dissolved in THF (15 mL) and TBAF·$3H_2O$ (0.021 mol, 6.7 mg) was added. After reaction and solvent removal, the brown solid obtained was purified by chromatography on silica gel using $CHCl_3$/MeOH (20:1) as eluent. **A1.1** was obtained as a white solid (7.95 g, 98%).

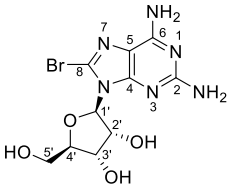
^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm): 7.04 (s (broad), 2H, NH_2), 6.06 (s (broad), 2H, NH_2), 6.00 (d, J = 2.4 Hz, 1H, $H^{1'}$), 5.42 (dd, J = 6.2, J' = 2.4 Hz, 1H, $H^{2'}$), 5.09 (dd, J = 6.2, J' = 3.4 Hz, 2H, $H^{3'}$), 4.83 (s, 1H, -CC*H*), 4.11 (dt, J = 5.8, J' = 3.3 Hz, 2H, $H^{4'}$), 3.64 – 3.45 (m, 2H, CH_2-OH), 1.53 (s, 3H, -OC(CH_3)), 1.32 (s, 3H, -OC(CH_3)).



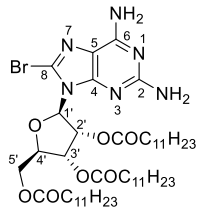
A1. A1.1 (0.86 mmol, 300 mg) and imidazole (1.72 mmol, 117 mg) were placed into a dried bottomed flask. Then DMF (11 mL) and TBDMSiCl (1.72 mmol, 258 mg) were added. The mixture was stirred for 12 hours at room temperature. The solvent was evaporated under reduced pressure, leaving a brown oil that was purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1). **A1** was obtained as a white solid (5.31 g, 43%).

^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.98 (s (broad), 2H, NH_2), 6.10 (s (broad), 2H, NH_2), 6.03 (s, 1H, $H^{1'}$), 5.50 (dd, J = 6.2, J' = 1.5 Hz, 1H, $H^{2'}$), 5.12 (dd, J = 6.3, J' = 3.5 Hz, 1H, $H^{3'}$), 4.80 (s, 1H, -CC*H*), 4.09 (dd, J = 7.3, J' = 5.7 Hz, 1H, $H^{4'}$), 3.72 – 3.69 (m, 2H, $H^{5'}$), 1.67 (s, 3H, -OC(CH_3)), 1.48 (s, 3H, -OC(CH_3)), 0.92 (s, 9H, -Si(CH_3)$_3$-C(CH_3)$_2$), 0.01 (d, J = 4.8 Hz, 6H, -Si(CH_3)$_3$-C(CH_3)$_2$).

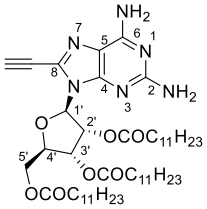
^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 160.4, 156.0, 150.8, 130.7, 114.3, 113.9, 89.9, 87.7, 83.3, 82.9, 82.4, 72.9, 63.3, 27.2, 25.9, 25.6, 18.4, -5.3, -5.4.
MS (FAB+): 461.2 $[M+H]^+$.



A2.2. The synthesis of **G1.4** was performed according to a published procedure.[145] The pristine 2-aminoadenosine (54.14 mmol, 15 gr) was suspended in H_2O. Afterwards, a bromine (164 mmol, 13.16 gr) solution in H_2O (422 mL) was added stepwise. After reaction completion, Na_2SO_3 was added and the PH was set at 6-7 to precipitate **A2.2.** The resulting crude was filtered and washed with cold water affording **A2.2** as a white solid (17.8 gr, 93%).
^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 7.08 (s (broad), 2H, NH_2), 5.79 (s (broad), 2H, NH_2), 6.69 (d, J = 5.5 Hz, 1H, $H^{1'}$), 5.65 – 5.61 (m, 1H, $H^{2'}$), 5.48 – 5.35 (m, 1H, $H^{3'}$),5.16 – 5.01 (m, 3H, $H^{4'}$, $H^{5'}$).

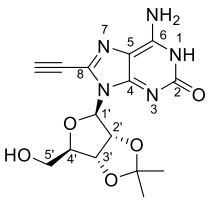


A2.1. In a 500-mL round-bottomed flask, equipped with a magnetic stirrer, **A2.2**[149] (36.0 mmol, 13 g) and DMAP (22 mmol, 2.67 mg) were placed. Dry THF/DMF (2:1) (250 mL) was added and the mixture was stirred at room temperature under argon until the solid was dissolved. Then, NEt_3 (16 mmol, 22.78 mL) and lauroyl chloride (112 mmol, 26 mL) were added. The resulting mixture was stirred at 130 ^{o}C until **A2.2** was consumed. Afterwards, MeOH (7 mL) was added and the mixture was stirred during 15 minutes. The solvent was then eliminated under reduced pressure and the solid was directly purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (80:1). **A2.1** was obtained as a pale solid (19.98 g, 59%).
^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 6.28 (d, J = 5.5 Hz, 1H, $H^{1'}$), 6.20 – 6.04 (m, 1H, $H^{2'}$), 5.94 (s, 1H, $H^{3'}$), 5.63 (s (broad), 2H, NH_2), 4.95 (s (broad), 2H, NH_2), 4.62 – 4.23 (m, 3H, $H^{4'}$, $H^{5'}$), 2.30 (dt, J = 32.2, J' = 8.4 Hz, 6H, -OCOCH_2-), 1.25 (m, 54H, -CH_2-), 0.86 (t, 9H, -CH_3).
^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 178.0, 173.5, 172.3, 162.4, 159.2, 154.7, 152.1, 122.5, 114.4, 88.5, 79.5, 72.3, 70.3, 62.7, 36.4, 33.8, 31.8, 31.3, 29.5, 29.4, 29.2, 29.1, 29.0, 24.7, 24.6, 22.6, 14.0.
MS (MALDI): Calculated for $C_{46}H_{79}BrN_6O_7$: 907.5272 $[M+H]^+$. Found: 929.5086 $[M+Na]^+$. Matrix: DCTB



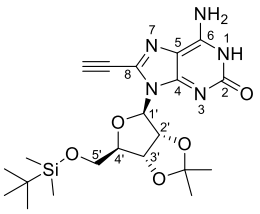
A2. **A2**was prepared following *Standard Procedure A*. **A2.1** (22.0 mmol, 19.8 g), $PdCl_2(PPh_3)_2$ (0.4 mmol, 306 mg), CuI (0.2 mmol, 42.0 mg), TMSA (28 mmol, 4.34 g) were mixed in the THF/NEt_3 solvent (20 mL). The mixture was then stirred at 40 ^{o}C during 24 h. After removal of the solvent, a brown oil was obtained that was used in the following reaction step without previous purification. Then, the brown oil was dissolved in THF (15 mL) and TBAF·$3H_2O$ (0.021 mol, 6.7 mg) was added. After reaction and solvent removal, the brown oil obtained was purified by chromatography on silica gel using $CHCl_3$/MeOH (60:1) as eluent. **A2** was obtained as a brown oil-solid (14.63 g, 78%).
^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 6.22 (s, 1H, $H^{1'}$), 6.11 (d, J = 6.3 Hz, 2H, $H^{2'}$, $H^{3'}$), 5.56 (s (broad), 2H, NH_2), 4.96 (s (broad), 2H, NH_2), 4.53 (q, J = 6.8 Hz, 1H, $H^{4'}$), 4.36 (d, J = 9.2 Hz, 2H, $H^{5'}$), 3.46 (s, 1H, -CCH), 2.33 (ddt, J = 23.4, J' = 15.5, J'' = 8.0 Hz, 6H, -OCOCH_2-), 1.28 (d, J = 15.3 Hz, 54H, -CH_2-), 0.88 (d, J = 7.1 Hz, 9H, -CH_3).
^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 173.5, 172.1, 160.0, 155.8, 150.9, 132.0, 128.3, 103.8, 87.4, 83.6, 79.6, 72.3, 70.4, 62.8, 47.0, 33.8, 31.9, 29.6, 24.8, 22.6, 14.0.
MS (MALDI): Calculated for $C_{48}H_{80}N_6O_7$: 854.2110 $[M+H]^+$. Found: 875.5975 $[M+Na]^+$. Matrix: DCTB

<u>Synthesis of the *iso*guanosine derivatives.</u>



iG1.2. iG1.2 was synthesized according to a literature procedure[146] that was adapted to our compound. **A1.1** (6.73 mmol, 2.82 g), $NaNO_2$ (20.9 mmol, 1.85 g) and AcOH (46.1 mmol, 2.83 g) were dissolved in a H_2O/THF (1:1) mixture (30 mL). After stirring at 50 ^{o}C for 2 hours, the reaction was cooled down to room temperature and the solvent was evaporated. The crude product was purified by column chromatography using $CHCl_3$/MeOH (20:1) as the eluent. **iG1.1** was obtained as a light orange solid (1.64 g, 70%).
^{1}H-NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.57 (s(broad), 1H, NH), 6.21 (s(broad), 1H, NH_2), 5.95 (d, 1H, J = 4.5 Hz, $H^{1'}$), 5.24 (s(broad), 1H, NH_2), 4.91 (dd, 1H, J = 5.8, J' = 1.8 Hz, $H^{2'}$), 4.28 (s, 1H, $H^{3'}$), 4.00 (s, 1H, -

CC*H*), 3.75 (d, 1H, *J* = 12.5 Hz, $H^{4'}$), 3.63 (d, 2H, *J* = 10.5 Hz, $H^{5'}$), 1.52 (s, 3H, -OC(CH_3)), 1.1.28 (s, 3H, -OC(CH_3)).
MS (FAB+): 348.1 $[M+H]^+$.



iG1. Into a 100 mL bottomed flask **iG1.1** (5.15 mmol, 2.16 g), imidazole (10.3 mmol, 71 mg) and *tert*-butyldimethylchlorosilane (10.3 mmol, 1.55 g) were dissolved in dry DMF (30 mL). The reaction was stirred for 2 hours at room temperature and then concentrated under vacuum. The residue was dissolved in $CHCl_3$ and washed with water (2 x 50 mL). The product was purified by column chromatography eluted with $CHCl_3$/ MeOH (30:1). An orange solid was obtained (2.02 g, 85%).
^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 10.50 (s (broad), 1H, N*H*), 7.43 (s(broad), 2H, NH_2), 6.01 (d, 1H, *J* = 2.1 Hz, $H^{1'}$), 5.54 (d, 1H, *J* = 6.6 Hz, $H^{2'}$), 4.95 (dd, 1H, *J* = 6.4, *J'* = 3.5 Hz, $H^{3'}$), 4.11 – 4.05 (m, 1H, $H^{4'}$), 3.80 – 3.66 (m, 3H, $H^{5'}$, -CC*H*), 1.48 (s, 3H, -OC(CH_3)), 1.27 (s, 3H, -OC(CH_3)), 0.76 (s, 9H, -Si(CH_3)$_2$-C(CH_3)), 0.11 (s, 6H, -Si(CH_3)$_2$-C(CH_3)).
^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 155.5, 113.1, 108.8, 89.0, 87.4, 84.5, 82.2, 81.7, 72.5, 63.3, 26.8, 25.5, 25.1, 17.8, 0.
MS (FAB+): 462.2 $[M+H]^+$.

Chapter 2

Evaluation of Dimerization and Association Constant Between Lipophilic Mononucleosides

In this Chapter we will study de H-bonding dimerization and hetero-association processes between complementary units (**G-C**, **iG-iC**, **G-iC**, **iG-C** and **A-U**) by both ^{1}H-NMR and absorption spectroscopic measurements in four different solvents: Toluene, $CHCl_3$ (or $CDCl_3$):CCl_4 (2:3), $CHCl_3$, THF and DMF. Then, we will analyze the binding isotherms by adequate fitting programs in order to obtain the relevant association constants.[150] Guanine-cytosine and adenine or 2-aminoadenineuracile binding has already studied by a number of authors,[36] so this work offers new quantitative data on their association constants studied and analyzed by diverse methods. However, to the best of our knowledge, no data has been reported so far on the association between *iso*guanine and *iso*cytosine in organic solvents, or on the interactions between these non-natural bases and cytosine or guanine.

All the spectroscopic measurements were carried out by Jorge Camacho García and form part of his doctoral thesis that was developed at the same time.

All of these data can be found in the Supporting Information of our paper: *Org.Biomol. Chem.* **2015,** *13,* 4506-4513 (Ref: 151)

2.1 Results and Discussion.

2.1.1 Target Molecules.

The six ethynyl-substituted nucleobases synthesized in the previous Chapter: **G1**, **C1**, **iG1**, **iC1**, **A1** and **U1**, constitute a new relevant collection of synthetic intermediates for supramolecular chemistry. The rich and useful reactivity of the terminal triple bond, either through Sonogashira couplings or "click" cycloaddition reactions make these compounds convenient synthons for the preparation of complex self-assembled systems. In this Chapter, we wanted to couple these bases to a very simple π-conjugated oligophenylene-ethynylene moiety, resembling as much as possible the π-conjugated structure of our monomers (Figure2.1), so as to study self-association and self-recognition processes via H-bonding.

[150] P. Thordarson, *Chem. Soc. Rev.* **2011**, *40*, 1305–1323
[151] J. Camacho-García, C. Montoro-García, A. M. López-Pérez, N. Bilbao, S. Romero-Pérez, D. González-Rodríguez, *Org. Biomol. Chem.* **2015**, *13*, 4506–4513.

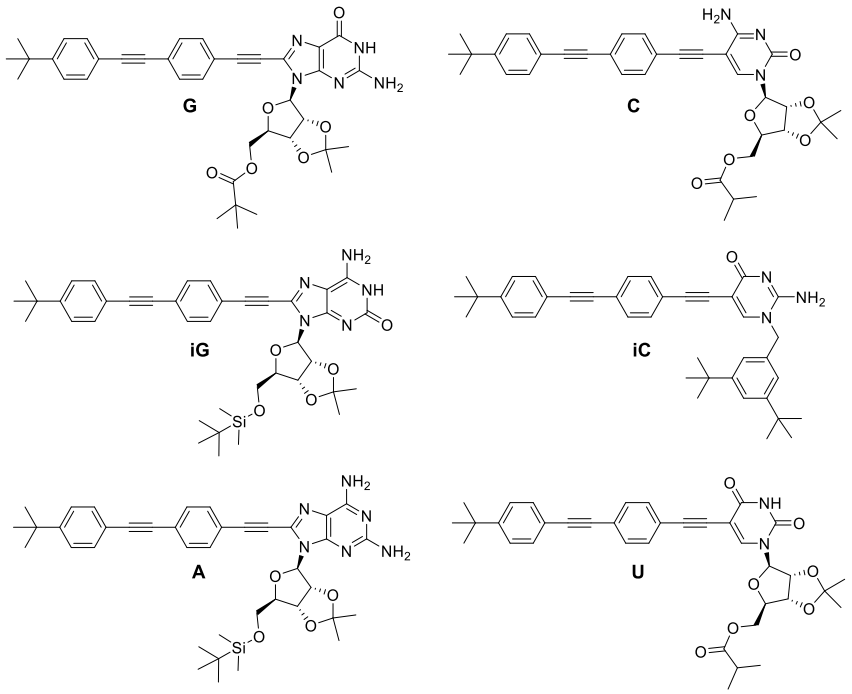
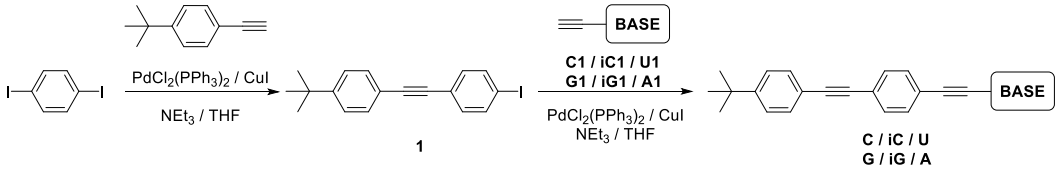


Figure 2.1. Structure of lipophilic nucleosides **G**, **C**, **iG**, **iC**, **A** and **U**.

2.1.2 Synthesis of Lipophilic Nucleosides.

For the synthesis of these π-extended nucleosides (**G**, **C**, **iG**, **iC**, **A** and **U**, see Scheme 2.1), iodoarene **1** was prepared first by Pd-catalyzed coupling between 4-*tert*butylphenylacetylene and 1,4-diiodobenzene, which was used in excess in order to maximize the yield of monosubstituted product. The common intermediate **1**[152] was then coupled with the corresponding ethynyl-terminated base (**G1**, **C1**, **iG1**, **iC1**, **A1** and **U1**), to produce the final compounds, which were purified and characterized by ^{1}H-NMR, ^{13}C-NMR, UV-vis, MS and HR-MS techniques. The synthetic details for the preparation of each derivative can be found at the Experimental Section (**2.4.1**) of this Chapter.



Scheme 2.1. Synthesis of lipophilic nucleosides **G**, **C**, **iG**, **iC**, **A** and **U**.

2.1.3 Evaluation of Dimerization and Association Constants.

In this Chapter we were particularly interested in assessing and comparing different experimental techniques and data analysis methods in order to determine equilibrium constants as accurately as possible. For these supramolecular equilibria, ^{1}H-NMR, absorption and emission spectroscopy techniques were considered. However, we found the last

[152] F. A. Murphy and S. M. Draper, *J. Org. Chem*. **2010**, *75*, 1862–1870.

technique not very reliable or practical due to the need to normalize each spectra by the total amount of absorbed light during the dilution experiment and due to the fact that both nucleosides absorbed and emitted in approximately the same regions (purine nucleosides's absorption and emission features were only slightly red-shifted with respect to those of pyrimidine nucleosides). In addition, emission spectroscopy is commonly used to calculate association constant in the order of 10^8 M^{-1} or even more and here we expect values in the order of 1-10^5 M^{-1}. Hence, only ^{1}H-NMR and UV-vis dilution and titration experiments were carried out.

The chemical shift or absorbance data as a function of concentration were then analyzed by different methods. For ^{1}H-NMR data, the software *Equilibria*[153] was found to be particularly handy and useful for these simple dimerization and 1:1 binding models.[154] Besides, the 1:1 binding can be fitted considering as well the possibility of host dimerization (see below). It is, however, a program for local analysis, meaning that the shift experienced by each 1H probe is analyzed independently. Some 1H-NMR host-guest binding data was also fitted with the Matlab® scripts developed by P. Thordarson,[150] that offers the possibility of global fitting. This means that several 1H probes can be fitted simultaneously, thus enhancing the quality of the fitting procedure. For the analysis of the optical absorbance data, however, the software *ReactLab*™ *EQUILIBRIA*[155] was the most appropriate one, since the whole spectra are globally fitted and both host and guest nucleoside dimerizations can be included in the fitting as competitive processes to the binding between nucleobase pairs.

Dimerization Constants. (K_{dim})

Before studying the binding events between complementary bases, we were interested to ascertain the extent of H-bonding aggregation in each final π-conjugated nucleoside. Since we did not expect significantly strong self-association for any of the **G**, **C**, **iG**, **iC**, **A** or **U** products (i.e. $K_{dim} \leq 10^3$ M^{-1}), we devised a set of dilution experiments that were adjusted to a simple dimerization model. The formation of higher order aggregates was thus neglected at low concentrations.

The system equilibrium, the corresponding dimerization constant equation, and the mass balance are:

$$M + M \leftrightarrow M_2 \qquad (1)$$

$$K_{dim} = [M_2]/[M]^2 \qquad (2)$$

$$[M]_0 = 2\,[M_2] + [M] \qquad (3)$$

where M is the corresponding π-conjugated mononucleoside. Hence, [M] can be expressed as a function of K_{dim} and $[M]_0$ as follows:

$$[M] = (-1 + (1 + (8\,K_{dim}\,[M]_0)^{1/2})/4K_{dim} \qquad (4)$$

The chemical shifts for the relevant nuclei (δ_{obs}) or absorbance values at a given wavelength (A_{obs}) are described as a weighted average of the individual species:

[153] The *Equilibria* program was developed by Christopher Marjo, Mark Wainwright Analytical Centre, University of New South Wales, Sydney, Australia. http://www.sseau.unsw.edu.au/.
[154] P. G. Young and K. A. Jolliffe, *Org. Biomol. Chem.* **2012**, *10*, 2664–2672.
[155] ReactLab™ EQUILIBRIA. Jplus Consulting Pty Ltd.

$$\delta_{obs} = \delta_M ([M]/[M]_0) + \delta_{M2} (2[M_2]/[M]_0) \quad (5)$$

$$A_{obs} = A_M ([M]/[M]_0) + A_{M2} (2[M_2]/[M]_0) \quad (6)$$

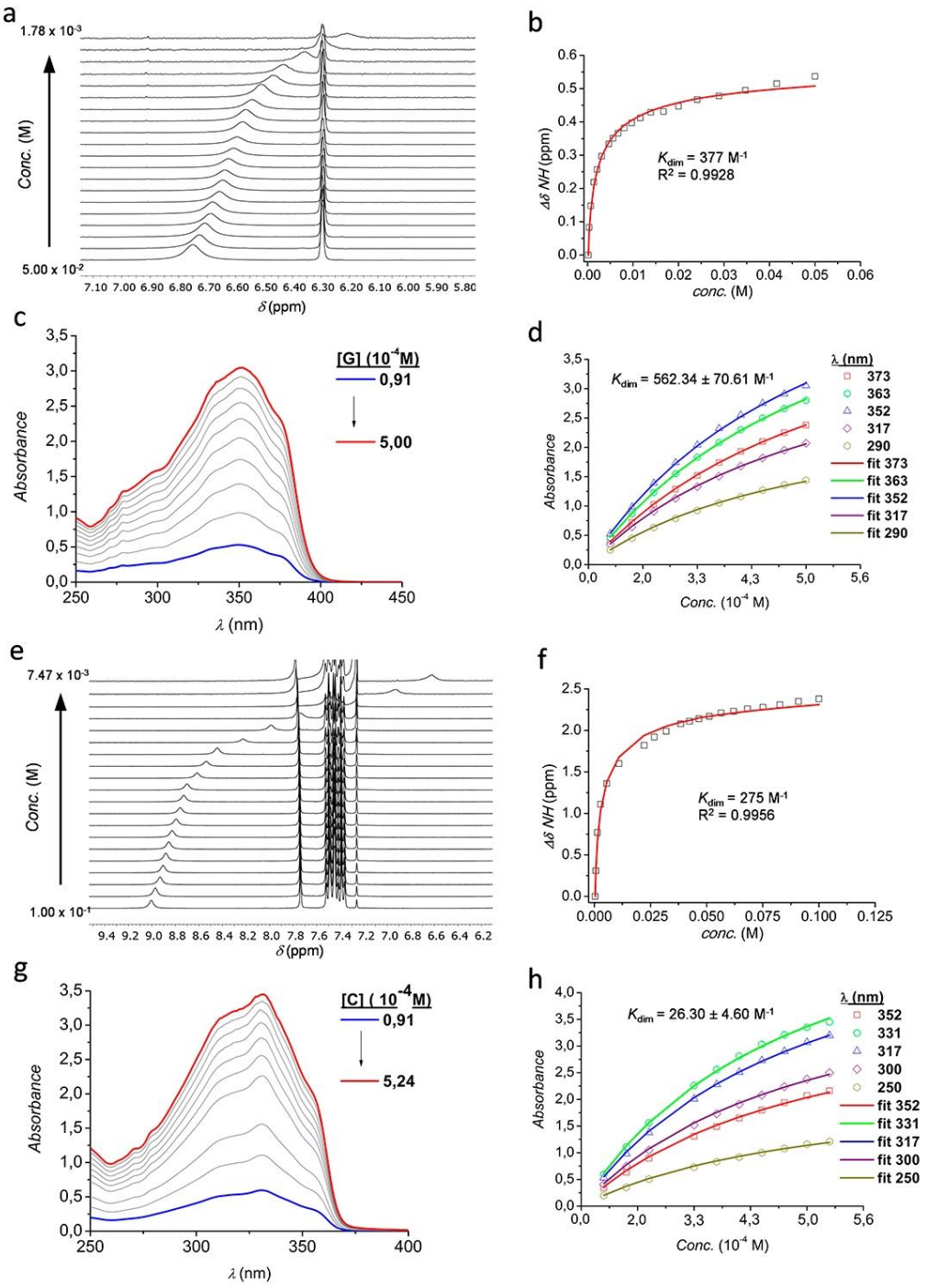


Figure 2.2. Selected regions of the ^{1}H NMR and UV-vis spectra as a function of **G** (**a,c**) and **C** (**e,g**) concentration showing the spectral changes occurring upon self-association. Right: fitting of the (**b,d**) **G** amide and (**f,h**) **C** amine proton chemical shift to a dimerization model.

The chemical shift and absorbance data as a function of total concentration were then analyzed by the different methods that have to deal with 3 unknown parameters: δ_M (or A_M), δ_{M2} (or A_{M2}) and K_{dim}. The dimerization constants obtained by these two techniques are displayed in Table 2.1. A representative example for both ^{1}H NMR and UV-vis concentration-dependent experiments and binding isotherms is given in Figure 2.2, but all of them can be found in the Supporting Information of our published work.[151,156]

Due to the low degree of self-association of our lipophilic nucleosides in chloroform, the equilibrium constants obtained from ^{1}H NMR experiments were found to be, in general, more reliable than those derived from absorption experiments. In the latter, a lower concentration must be used, which produced binding isotherms that mainly covered a region of low probability of binding (p).[1] That is especially evident in the case of the nucleosides with lower dimerization constants (i.e. **U** and **A**). The dimerization constants were also evaluated in the less polar 2:3 *v/v* $CHCl_3$ (or $CDCl_3$):CCl_4 solvent system, which produced a higher degree of self-association. Nevertheless, the dimerization constants obtained by both techniques are in the same order of magnitude and consistent with literature values.[36]

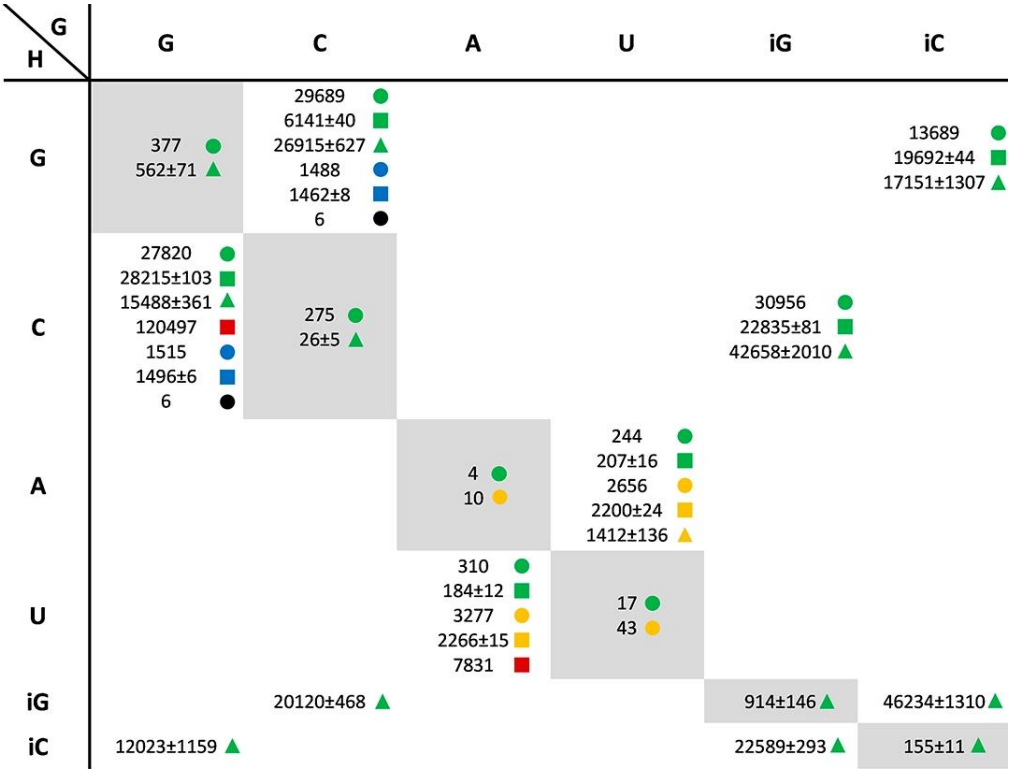


Table 2.1 Dimerization (K_{dim}) and association (K_a) constants calculated by ^{1}H NMR or UV-vis titration experiments with the different lipophilic nucleosides prepared in this Chapter.
Color code: **$CHCl_3$** (or **$CDCl_3$**). In **$CHCl_3$** (or **$CDCl_3$**):**CCl_4** (2:3). In **toluene-d_8**. In **THF-d_8**. In **DMF-d_7**.
Shape code: *Circle*: NMR data fitted with *Equilibria*[153] considering host dimerization. *Square*: NMR data fitted with the Matlab® scripts developed by P. Thordarson.[152] *Triangle*: UV-vis absorption data fitted with *ReactLab*™ EQUILIBRIA.[155]

[156] C. Montoro-García, J. Camacho-García, A. M. López-Pérez, N. Bilbao, S. Romero-Pérez, M. J. Mayoral, D. González-Rodríguez, *Angew. Chem. Int. Ed.* **2015**, *54*, 6780–6784 (VIP Paper).

In the case of the novel **iG** and **iC** nucleobases, dimerization constants were found to be relatively high, in the same order to those found for **G**. As a matter of fact, these products were not extraordinarily soluble in chloroform and their ^{1}H NMR spectra showed rather broad NH/NH_2 resonance peaks, which complicated the analysis by this technique at high concentrations. Dimerization constants were instead determined from UV-vis dilution measurements in this case.

On the other hand, in polar solvents like THF and DMF that compete strongly for H-bonding, self-association processes between nucleobases were considered negligible and were not introduced in the subsequent analysis of the binding constants between complementary nucleoside pairs.

2.1.4 Association Constants Between Complementary Nucleosides.

Next, we performed titration experiments between complementary bases in order to determine their association constants (K_a). Increasing amounts of a solution of the guest nucleoside (G) were added over a solution of the complementary host nucleoside (H). The guest solutions contained as well the host nucleoside, so that [H] was not altered during titration. We arbitrarily assigned the purines (**G**, **iG**, and **A**) as the hosts and the pyrimidine nucleosides (**C**, **iC**, and **U**) as guests, although we also performed and analyzed the opposite titration (see Table 2.1).

Only a 1:1 binding model was considered. In this case, the system equilibrium, the corresponding association constant equation, and the mass balance are:

$$H + G \leftrightarrow HG \quad (7)$$

$$K_a = [HG]/[H][G] \quad (8)$$

$$[G]_0 = [G] + [HG] + [G_2] \quad (9)$$

$$[H]_0 = [H] + [HG] + [H_2] \quad (10)$$

Notice that in the mass balance we also include the possibility of host and, in the absorption measurements, guest dimerization as a competitive reaction to host-guest binding. Hence, equations (1) and (2) were also considered in the analyses. The NMR probe on the host nucleoside has 3 chemical shifts corresponding to the species in solution: the unbound chemical shift (δ_H) the chemical shift of the complex with the guest (δ_{HG}) and the chemical shift the host dimer (δ_{H2}). The observed chemical shift will be a mixture of the 3 shifts according the mole fraction of each species present and can be calculated as:

$$\delta_{obs} = (\delta_H[H] + \delta_{HG}[HG] + \delta_{H2}[H_2]) / ([H] + [HG] + 2[H_2]) \quad (11)$$

On the other hand, in the UV-vis titration experiments, the normalized absorbance value at a given wavelength (A_{obs}) can be described as:

$$A_{obs} = (A_H[H] + A_{HG}[HG] + A_{H2}[H_2] + A_{G2}[G_2]) / ([H] + [HG] + 2[H_2] + 2[G_2]) \quad (12)$$

A typical set of titration experiments is shown in Figure 2.3 The chemical shift and absorbance data as a function of guest concentration were then analyzed. As stated above, NMR titrations were fitted with the software *Equilibria*,[154] which offers the possibility of including host dimerization in the analysis or, in some cases, with the Matlab® global fitting scripts developed by P. Thordarson,[150] which results in lower standard errors. UV-vis titrations

were instead fitted with the software *ReactLab*™ *EQUILIBRIA*[155] that includes both the whole spectra, as well as host and guest dimerization constants in the analysis. The binding constants obtained are displayed in Table 2.1.

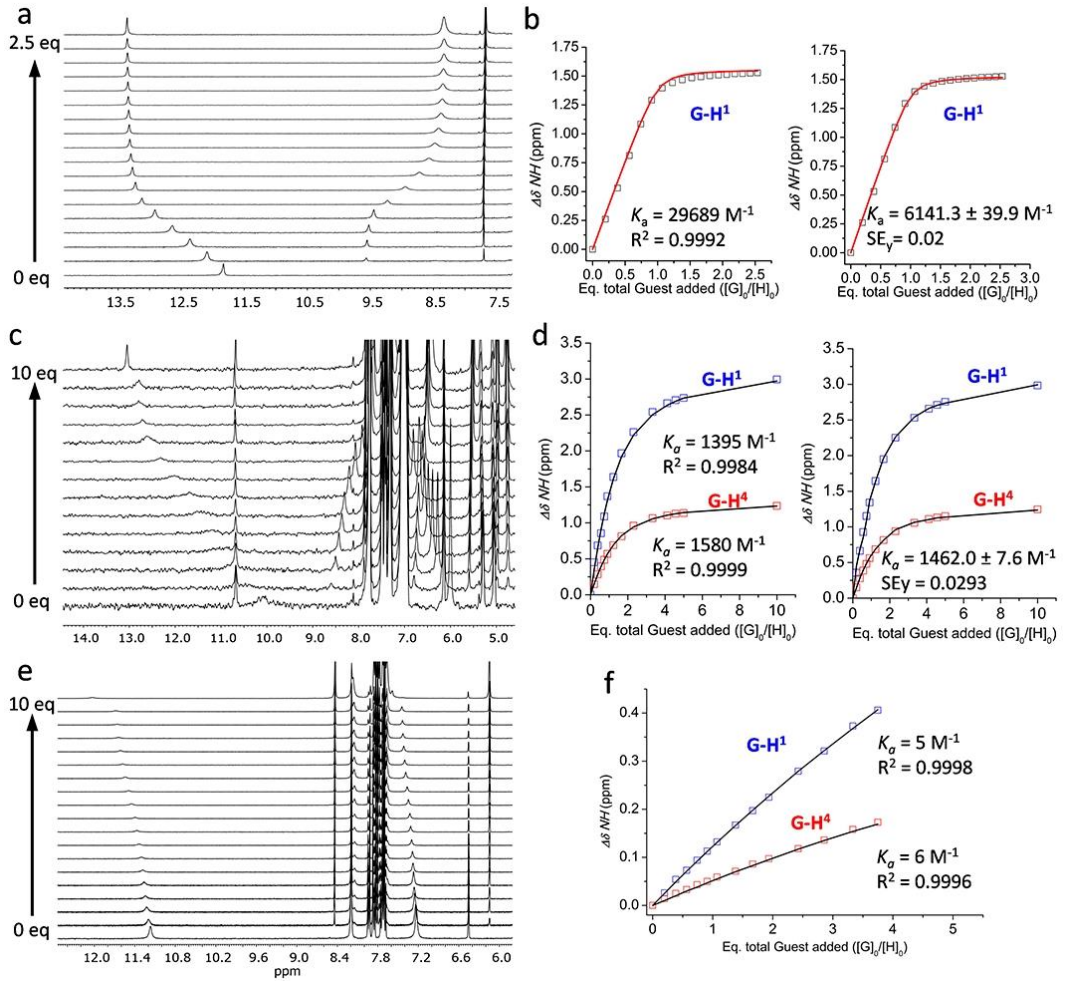


Figure 2.3. Selected regions of the ^{1}H NMR spectra obtained during the titration experiment of **G** (host) as a function of **C** (guest) concentration showing the spectral changes occurring upon association in $CDCl_3$ (**a**), THF-d_8 (**c**) and DMF-d_7 (**e**). Right: fitting of the (**b**,**d**,**f**) ^{1}H chemical shift to a 1:1 binding model.

Despite NMR and UV-vis titration experiments were performed employing two different techniques, concentration ranges and fitting programs, the 1:1 binding constants derived from these experiments are in reasonable agreement. In addition, the values obtained in the experiments where purines were considered as the hosts and pyrimidines as the guests are related to those acquired when the opposite titration order was applied (pyrimidines as hosts and purines as guests).

In the case of **G**-**C** or **A**-**U** base pairs, the association constants derived in this work match the values obtained by other authors (see Table 2.1).[36] Nucleosides **A** and **U**, having a symmetric *ADA-DAD* H-bonding pattern (Figure 4), associate in $CHCl_3$ (or $CDCl_3$) with K_a = 1.8–3.1 x 10^2 M^{-1}. In the less polar 2:3 *v/v* $CHCl_3$ (or $CDCl_3$):CCl_4 solvent mixture the association constants were found to increase by an order of magnitude, reaching K_a = 1.4–3.2 x 10^3 M^{-1}.

In the same line, the association constants calculated in toluene-d_8 (K_a = 7.8 x 10^3 M^{-1}) revealed a stronger **A**-**U** association than in $CDCl_3$.

On the contrary, the binding constants calculated for the asymmetric *DDA-AAD* H-bonding patterns (Figure 2.4), found in **G**-**C**, **G**-**iC**, **iG**-**C** and **iG**-**iC** base pairs, are in the order of 10^4 M^{-1}. For instance, the binding constants obtained in $CHCl_3$ (or $CDCl_3$) amount to: K_a = 1.2–2.0 x 10^4 M^{-1} (**G**-**iC**), K_a = 1.5–3.0 x 10^4 M^{-1} (**G**-**C**), K_a = 2.0–4.3 x 10^4 M^{-1} (**iG**-**C**), and K_a = 2.2–4.7 x 10^4 M^{-1} (**iG**-**iC**). It should be remarked that the **iG**-**C** and **G**-**iC** association, comparable in strength to **G**-**C** and **iG**-**iC** as demonstrated in this work, occurs *via* a reverse Watson-Crick H-bonding interaction pattern, as it is shown in Figure 2.4. The low value obtained in the case of the **G**(host) + **C** (guest) titration analyzed by Matlab scripts (see Table 2.1) is probably due to the fact that this fitting method does not consider host (**G**) dimerization. In fact, when the same data was analyzed with *Equilibria* ignoring **G** dimerization, a value of K_a = 6.6 x 10^3 M^{-1} was calculated. To conclude, the data obtained from the different fittings in THF-d_8 (K_a = 1.4 x 10^3 M^{-1} (**G**-**C**)) and DMF-d_7 (K_a = 6 M^{-1} (**G**-**C**)) reveals a notable decrease of the association constant due to the high affinity of these solvents for H-bonding to the Watson-Crick sides.

Finally, in the less polar solvent, toluene, association constants between **G** and **C** nucleosides were calculated as K_a = 1.2 x 10^5 M^{-1}, although the experimental titration data should not be fully trusted in this case because: i) the binding constant is a bit too high to be determined within the NMR concentration range and, ii) the mononucleosides showed clear signs of self-aggregation in the ^{1}H NMR spectra and they are not fully soluble at concentrations above 10^{-2} M. Due to these two reasons, the titrations were carried out at the minimum concentration possible: 1.0 x 10^{-4} M for the stronger **G**-**C** pair and 5.0 x 10^{-4} M for the weaker **A**-**U** pair. Recent results in the group afforded similar binding constants between related complementary mononucleosides featuring donor-acceptor energy transfer moieties, as measured by optical spectroscopic techniques (K_a = 2.0 x 10^3 for the **A**-**U** and K_a = 3.0 x 10^5 for the **G**-**C** couple).

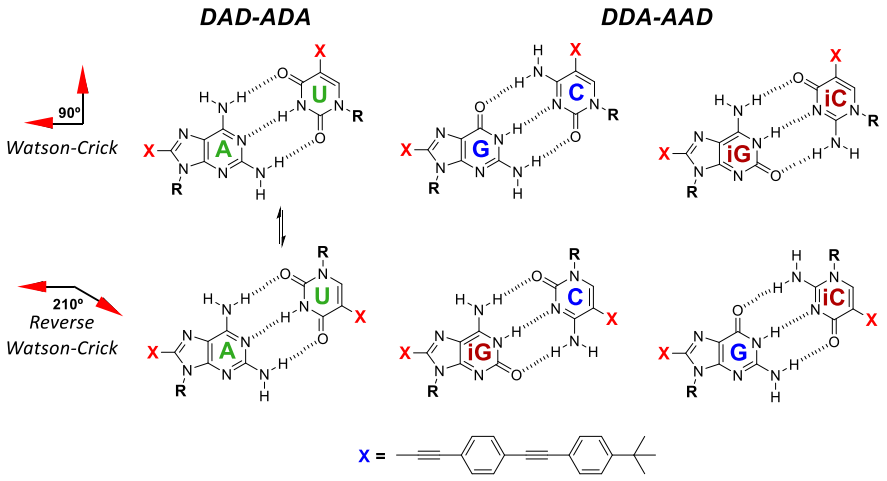


Figure 2.4. Triply H-bonded self-complementary base pair structures considered in this work.

2.2 Conclusions.

In this Chapter we have prepared a series of lipophilic nucleosides comprising natural and non-natural bases that are π-conjugated to a short oligophenylene-ethynylene fragment. These bases include guanosine, *iso*guanosine, 2-aminoadenosine as purine heterocycles, and cytidine, *iso*cytosine and uridine as complementary pyrimidine bases. The H-bonding dimerization and association processes between complementary bases were as well evaluated using different techniques (^{1}H NMR and absorption spectroscopies), solvents, concentration ranges and fitting programs.

Symmetric *ADA-DAD* H-bonding patterns (**A**-**U** base pairs) produce 1:1 binding constants in the order of 10^2-10^4 M^{-1}, whereas unsymmetric *DDA-AAD* H-bonding patterns (**G**-**C**, and the novel **G**-**iC**, **iG**-**C** and **iG**-**iC** base pairs) yield association constants in the order of 10-10^5 M^{-1}, depending on the polar nature of the solvent. Such increase in approximately two orders of magnitude in $CHCl_3$ when **G**-**C** and **A**-**U** associating pairs are compared is well-known in the literature and is caused by the establishment of stabilizing secondary H-bonding interactions in the *DDA-AAD* pairs, as explained in the Introduction of this Thesis.[36]

This work provides, to the best of our knowledge, the first association constant values between complementary nucleobases in diverse solvents like toluene, THF or DMF. It also includes K_a calculations for the non-natural **iG**-**iC** pair and all possible combinations of the purine-pyrimidine pairs between **G**, **C**, **iG** and **iC** bases, which reveal comparable K_a values as a consequence of the similar *DDA-AAD* triple H-bonding patterns. This fact suggests a lack of binding selectivity between these nucleobases and will be addressed again in *Chapter 4* through self-sorting experiments.

2.3 Experimental Section.

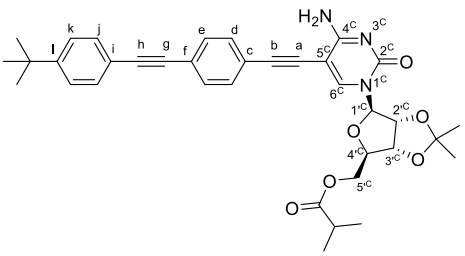
The General Methods detailed in the Experimental Section **1.4** of *Chapter 1* are also applicable here. The work described in this Chapter can also be found in the Supporting Information of our paper: Org. Biomol. Chem. **2015**, *13*, 4506–4513 (ref.XX).

Spectroscopy measurements. All the instruments are located at Universidad Autónoma de Madrid. **UV-Visible** experiments were conducted using a *JASCO V-660* apparatus. **Emission spectra** were recorded in a *JASCO FP-8600* equipment. **CD spectra** were recorded with a *JASCO* V-815 equipment. In all these three instruments the temperature was controlled using a *JASCO* Peltier thermostatted cell holder with a range of 263–383 K, adjustable temperature slope, and accuracy of ± 0.1 K.

2.3.1. Synthesis and Characterization.

The synthesis and characterization of the 5-/8-ethynilated nucleobases were described in the previous *Chapter 1* (section **1.4.**). The synthesis and characterization of iodoarene **1**[152] has been reported elsewhere.

Standard Procedure E for the Sonogashira coupling between the ethynyl-nucleobase and iodoarene 1. A dry THF/ NEt_3 (4:1) solvent mixture was subjected to deoxygenation by three freeze-pump-thaw cycles with argon. Then, this solvent was added over the system containing the corresponding ethynyl-substituted base (1.1 eq.), iodoarene **1**[4] (1 eq.), CuI (0.01 eq.) and $Pd(PPh_3)_2Cl_2$ (0.02 eq.). The reaction is stirred under argon at a given temperature and for a period of time (indicated in each case) until completion, which was monitored by TLC. Then, the mixture was filtrated over celite and the solvent evaporated under vacuum. The resulting crude product was purified by column chromatography (eluent indicated in each case).



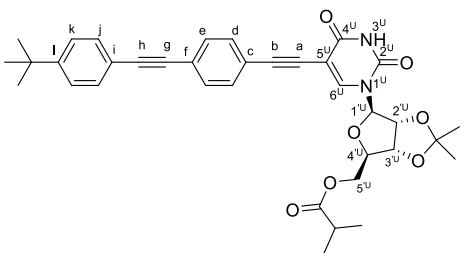
C. Lipophilic nucleoside **C** was prepared according to *Standard Procedure E*. **C1** (148.7 mg, 0.24 mmol), iodoarene **1**[152] (118 mg 0.3 mmol), $Pd(PPh_3)_2Cl_2$ (7.6 mg), CuI (0.6 mg) and THF/NEt_3 (2 mL) were mixed. The reaction was stirred during 12 h at 40 ^{o}C. The product was purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1), yielding **C** as a yellow solid (88 mg, 44%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 9.13 (s (broad), 1H, N*H*), 7.74 (s, 1H, H^{6C}), 7.57 – 7.33 (m, 8H, $H^{d, e, j, k}$), 6.00 (s (broad), 1H, NH_2), 5.73 (d, J = 1.7 Hz, 1H, $H^{1'C}$), 4.97 (d, J = 6.4, 1H, $H^{2'C}$), 4.80 (dd, J = 6.3, J' = 3.6 Hz, 1H, $H^{3'C}$), 4.53 – 4.18 (m, 3H, $H^{5'C}$), 2.58 (m, 1H, -OCOC*H*$(CH_3)_2$), 1.49 (s, 3H, -OC(CH_3)), 1.27 (s, 3H, -OC(CH_3)), 1.25 (s, 9H, -C(CH_3)), 1.15 (dd, J = 7.0, J' = 2.2, 6H, -OCOCH(CH_3)$_2$).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 176.7, 164.9, 154.0, 152.1, 145.2, 131.8, 131.5, 131.5, 125.6, 124.3, 121.5, 120.0, 114.3, 95.7, 95.4, 92.1, 91.6, 88.4, 85.9, 85.7, 81.2, 81.1, 77.6, 77.2, 76.8, 64.3, 35.0, 34.0, 31.3, 27.3, 25.4, 19.2, 19.0.

HRMS (FAB+): Calculated for $C_{36}H_{40}N_3O_6$: 610.2811 $[M+H]^+$. Found: 610.2902, $[M+H]^+$.

UV-Vis ($CHCl_3$) = λ_{max} = 331 nm, 335 (sh) nm.



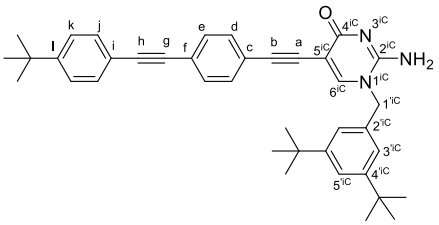
U. Lipophilic nucleoside **U** was prepared according to *Standard Procedure E*. Compound **U1** (0.41 mmol, 149 mg), iodoarene **1**[152] (0.33 mmol, 119 mg), $Pd(PPh_3)_2Cl_2$ (8 mg) and CuI (0.01 eq, 3.3 µmol, 1 mg) were dissolved in THF/NEt_3 4:1 (3 mL). The reaction was stirred during 12h at 40 °C. The crude material was purified by column chromatography using $CHCl_3$/ MeOH (30:1) as eluent. Recrystallization in MeOH yielded **U** as a yellow solid (168 mg, 84%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.92 (s (broad), 1H, CON*H*), 7.59 (s, 1H, H^{6C}), 7.42 – 7.28 (m, 8H, $H^{d, e, j, k}$), 5.74 (d, 1H, J = 2.2 Hz, $H^{1'C}$), 4.83 (dd, 1H, J = 6.4, J' = 2.3 Hz, $H^{2'C}$), 4.71 (dd, 1H, J = 6.4, J' = 3.8 Hz, $H^{3'C}$), 4.38 – 4.30 (m, 1H, $H^{4'C}$), 4.27 (d, 2H, J = 4.3 Hz, $H^{5'C}$), 2.54 (m, 1H, J = 7.0 Hz, -OCOC*H*$(CH_3)_2$), 1.51 (s, 3H, -OC(CH_3)), 1.28 (s, 3H, -OC(CH_3)), 1.24 (s, 9H, -C(CH_3)), 1.10 (d, 3H, J = 5.2 Hz, -OCOCH($CH_3)_2$), 1.07 (d, 3H, J = 5.1 Hz, -OCOCH($CH_3)_2$).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 176.5, 161.0, 151.9, 148.9, 143.3, 131.6, 131.5, 131.4, 125.4, 124.0, 121.9, 119.9, 114.9, 100.6, 93.9, 93.7, 91.8, 88.3, 85.0, 84.9, 81.3, 80.6, 63.7, 34.8, 33.9, 31.2, 27.2, 25.3, 19.1, 18.9.

HRMS (FAB+): Calculated for $C_{36}H_{38}N_2O_7$: 610.2679 $[M+H]^+$. Found: 610.2689 $[M+H]^+$.

UV-Vis ($CHCl_3$) = λ_{max} = 330 nm.



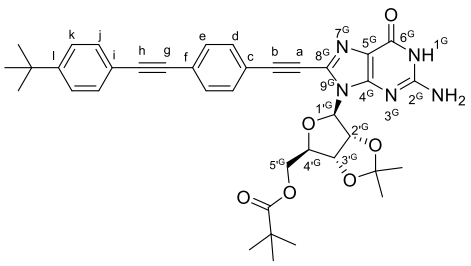
iC. Product **iC** was prepared according to *Standard Procedure E*. Iodoarene **1**[152] (0.33 mmol, 0.120 g), **iC1** (0.40 mmol, 0.143 g), $Pd(PPh_3)_2Cl_2$ (0.02 eq, 6.6 µmol, 4 mg) and CuI (0.01 eq, 3.32 µmol, 1 mg) were dissolved in NEt_3/THF (4:1) (5 mL). The reaction was stirred during 12h at 40° C The crude material was purified by chromatography column using $CHCl_3$/MeOH (30:1) as eluent. Recrystallization in MeCN yielded **iC** as a yellow solid (76 mg, 40%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.44 (s, 1H, H^{6iC}), 7.41 (s, 2H, H^k), 7.38 (d, 4H, J = 8.5 Hz, H^e, H^j), 7.30 (d, J = 8.0 Hz, 3H, $H^{5'iC}$, H^d), 6.95 (s, 2H, $H^{3'iC}$), 5.87 (s (broad), 2H, NH_2), 4.83 (s, 2H, $H^{1'iC}$), 1.25 (s, 9H, -C($CH_3)_3$), 1.25 (s, 18H, -C($CH_3)_3$).

^{13}C-NMR (75 MHz, DMSO-d_6) δ(ppm) = 167.6, 154.9, 152.2, 151.2, 147.4, 135.1, 132.0, 131.7, 131.6, 126.1, 123.4, 122.5, 121.9, 121.5, 119.6, 114.1, 102.9, 100.0, 91.8, 91.3, 88.8, 87.5, 53.8, 35.1, 35.0, 31.6, 31.3.

HRMS (FAB+): Calculated for $C_{39}H_{45}N_3O$: 570.3406 $[M+H]^+$. Found: 570.3499, $[M+H]^+$.

UV-Vis ($CHCl_3$) = λ_{max} = 382 nm, 408(sh) nm.



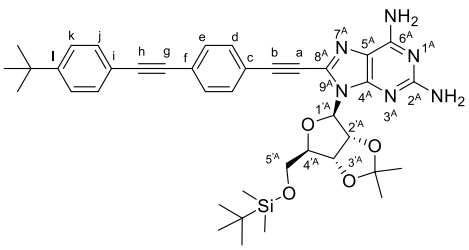
G. Lipophilic nucleoside **G** was prepared according to *Standard Procedure E*. **G1** (200 mg, 0.46 mmol), iodoarene **1**[152] (200 mg 0.56 mmol), $Pd(PPh_3)_2Cl_2$ (6 mg, 5 µmol), CuI (1 mg, 2.5 µmol) and THF/NEt_3 (5 mL), the reaction was stirred during 12 h at 40 °C. **G** was purified by chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1), obtaining a yellow solid (232 mg, 76%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 11.68 (s(broad), 1H, N*H*), 7.55 – 7.36 (m, 8H, $H^{d, e, j, k}$), 6.75 (s(broad), 2H, NH_2), 6.30 (dd, J = 6.1, J' = 1.7 Hz, 1H, $H^{1'G}$), 5.48 (dd, J = 6.3, J' = 3.5 Hz, 1H, $H^{2'G}$), 4.72 (dd, J = 11.1, J' = 7.2 Hz, 1H, $H^{3'G}$), 4.32 (dd, J = 6.7, J' = 3.5 Hz, 1H, $H^{4'G}$), 4.08 (dd, J = 11.1, J' = 6.4 Hz, 1H, $H^{5'G}$), 1.54 (s, 3H, -OC(CH_3)), 1.32 (s, 3H, -OC(CH_3)), 1.26 (s, 9H, -C($CH_3)_3$), 1.11 (s, 9H, -OCOC($CH_3)_3$).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 177.1, 155.9, 153.3, 151.1, 130.8, 130.7, 130.4, 124.6, 123.6, 19.6, 118.6, 112.9, 92.6, 91.5, 88.7, 87.3, 85.0, 83.3, 81.4, 77.6, 77.2, 76.7, 63.3, 40.0, 39.7, 39.4, 39.1, 38.8, 38.6, 37.7, 33.9, 30.2, 26.4, 26.2, 24.7.

HRMS (FAB+): Calculated for $C_{38}H_{42}N_5O_6$: 664.3057 $[M+H]^+$. Found: 664.2686 $[M+H]^+$.

UV-Vis ($CHCl_3$) = λ_{max} = 351 nm, 374 (sh) nm.



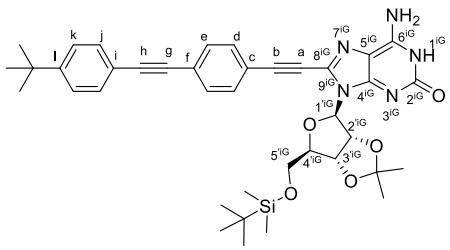
A. Lipophilic nucleoside **A** was prepared according to *Standard Procedure E*. **A1** (159 mg, 0.23 mmol), iodoarene **1**[152] (104 mg, 0.3 mmol), $Pd(PPh_3)_2Cl_2$ (6.7 mg, 6 µmol), CuI (2 mg, 3 µmol) and THF/NEt_3 (2 mL) were mixed. The reaction was stirred during 12 h at 40 °C. Compound **A** was purified by chromatography on silica gel using $CHCl_3$/MeOH (20:1) as eluent, obtaining a yellow solid (174 mg, 87%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.63 – 7.29 (m, 8H, $H^{d, e, j, k}$), 6.27 (d, J = 2.0 Hz, 1H, $H^{1'A}$), 5.80 (s (broad), 2H, NH_2), 5.67 (dd, J = 6.3, J' = 2.0 Hz, 1H, $H^{2'A}$), 5.09 (dd, J = 6.3, J' = 3.3 Hz, 1H, $H^{3'A}$), 4.81 (s (broad), 2H, NH_2), 4.27 (dd, J = 6.7, J' = 3.3 Hz, 1H, $H^{4'A}$), 3.77 (m, 2H, $H^{5'A}$), 1.63 (s, 3H, -OC(CH_3)), 1.42 (s, 3H, -OC(CH_3)), 1.33 (s, 9H, -C(CH_3)$_3$), 0.85 (s, 9H, -Si(CH$_3$)$_2$(CH_3)$_3$), 0.00 (d, J = 5.3 Hz, 6H, -Si(CH_3)$_2$(CH$_3$)$_3$).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 160.2, 155.9, 152.2, 151.2, 132.2, 132.0, 131.7, 131.6, 125.6, 125.1, 120.5, 119.9, 115.1, 114.0, 94.7, 92.6, 90.2, 88.4, 87.9, 83.2, 82.6, 80.0, 77.6, 77.2, 76.7, 63.6, 35.0, 31.3, 27.5, 26.0, 25.8, 18.5, -5.2, -5.3.

HRMS (FAB+): Calculated for $C_{39}H_{49}N_6O_4Si$: 693.3552 $[M+H]^+$. Found: 693.3596, $[M+H]^+$.

UV-Vis ($CHCl_3$) = λ_{max} = 357 nm, 384 (sh) nm.



iG. Product **iG** was prepared according to *Standard Procedure E*. Compound **iG1** (0.43 mmol, 200 mg), iodoarene **1**[152] (0.52 mmol, 0.187 g), $Pd(PPh_3)_2Cl_2$ (8.66 µmol, 6.07 mg) and CuI (4.33 µmol, 0.83 mg) were dissolved in THF/NEt_3 4:1 (3 mL). The reaction was stirred during 12h at 40° C. The crude material was purified by column chromatography using $CHCl_3$/MeOH (30:1) as eluent. Recrystallization in MeCN yielded iG as a yellow solid (255 mg, 85%).

^{1}H-NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.60 – 7.26 (m, 8H, $H^{d, e, j, k}$), 6.14 (s, 1H, $H^{1'iG}$), 5.46 (s(broad), 1H, NH_2), 4.98 (s, 1H, $H^{2'iG}$), 4.34 (s, 1H, $H^{3'iG}$), 4.08 – 3.72 (m, 2H, $H^{5'iG}$), 3.12 (d, 1H, J = 7.3 Hz, $H^{4'iG}$), 1.61 (s, 3H, -OC(CH_3)), 1.39 (s, 3H, -OC(CH_3)), 1.28 (s, 9H, -C(CH_3)$_3$), 0.83 (s, 9H, -Si(CH$_3$)$_2$(CH_3)$_3$), -0.00 (s, 6H, -Si(CH_3)$_2$(CH$_3$)$_3$).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 151.9, 137.3, 132.8, 131.6, 131.5, 131.2, 131.1, 125.3, 124.8, 120.3, 120.1, 119.4, 113.7, 100.0, 92.4, 88.0, 77.7, 63.7, 46.3, 34.7, 31.0, 27.2, 25.8, 25.4, 18.2, 8.7, -5.5, -5.5.

HRMS (FAB+): Calculated for $C_{39}H_{48}N_5O_5Si$: 694.3346 $[M+H]^+$. Found: 694.3435 $[M+H]^+$.

UV-Vis ($CHCl_3$) = λ_{max} = 361 nm.

In this moment, we though appropriate comment briefly the distribution of the ^{1}H-NMR ribose proton signals corresponding to the monomers studied in this work, which are shown in Figure 2.5. It is interesting to note that $H^{1'}$ (6.3 ~ ppm) is more down-shifted in purines than pyrimidines. The reason of this difference lies on the proximity of this position to the Nitrogen N(9), N(7) and N(3) placed in the purine ring and the two nearby Oxygen atoms placed in the ribose moiety which modify the electronic properties of that position. Afterwards, following a similar distribution influenced by Nitrogen N(9) and the Oxygen of the ribose, the $H^{2'}$, $H^{3'}$ and $H^{4'}$ protons display up-shifted signals (5.6 – 4.2ppm) in the same order in both families. Finally, the more down-shifted signal correspond to the 5' ribose position, which is highly influenced by the substitution chosen for that position. Thus, this signal $H^{5'}$ can be found either in the same shift as $H^{4'}$, splitted in two signals or even forming a lone signal.

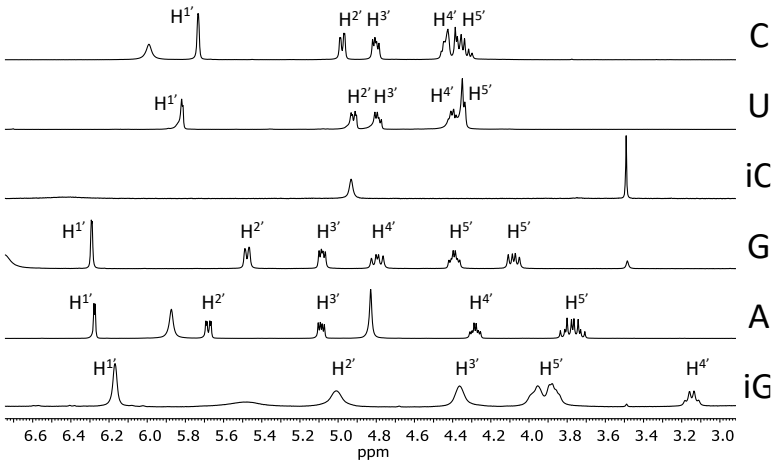


Figure 2.5. ^{1}H NMR spectra of the pyrimidines (**C**, **U**, **iC**) and purines (**G**, **A**, **iG**)

2.3.2 NMR and UV-vis Dilution and Titration Experiments.

NMR dilutions and titrations were carried out in in 5 mm NMR tubes using DMF-d_7, THF-d_8, $CDCl_3$ or 2:3 *v/v* $CHCl_3$-CCl_4. Deuterated solvents were purchased from Aldrich in ampoules and used as received. Residual $CHCl_3$ was used as the internal references (7.26 ppm), respectively. UV-vis dilutions and titrations were carried out in $CHCl_3$ or 2:3 *v/v* $CHCl_3$-CCl_4 (Alfa Aesar, Spectrophotometric Grade). The experiments were performed in 1 cm or 1 mm path length quartz cuvettes. Volumes were added using Hamilton microsyringes. UV-vis absorbances were kept within the 0.2-3.5 range. Temperature control was set at 298 K in all cases.

Dilution experiments were carried out by successive injections of a stock solution of the corresponding nucleoside monomer into clean solvent, thus increasing the concentration along the experiment. We found this method more practical and reliable than performing successive dilutions of the concentrated starting sample. The full ^{1}H NMR/UV-vis spectra were recorded over at least 15 concentrations, considering, as far as possible, that most of them should yield chemical shift/absorbance data within the 20-80% saturation range. Hence, the concentration range targeted depended on the dimerization constant expected for each nucleobase. Each dilution experiment was repeated at least twice.

Titration experiments were performed as follows. A sample of the host nucleoside was dissolved in the appropriate solvent, whose concentration, indicated in each experiment below, varied depending on the technique employed (^{1}H NMR or UV-Vis) and the expected magnitude of the association constant. A portion of this solution was used as the host sample, and the remainder was used to dissolve the sample of the guest, so that the host concentration remained constant throughout the titration. Successive aliquots of the guest solution, typically 10-20 times more concentrated, were added to the host sample, and the whole ^{1}H NMR / UV-vis spectra were recorded after each of the 15-20 guest additions. Again, in order to cover as much as possible the 20%-80% probability of binding range, the initial host concentration and the number of guest equivalents targeted was lower or higher as a function of the expected association constant between complementary bases. Each titration experiment was repeated at least twice.

a) *Equilibria* (^{1}H NMR).

The *Equilibria* program is a software package developed by Christopher Marjo, Mark Wainwright Analytical Centre, University of New South Wales, Sydney, Australia (http://www.sseau.unsw.edu.au/). It has been written using C++, and the Microsoft® Foundation Classes.

The NMR models for i) dimerization and ii) the formation of 1:1 HG complex + H_2 dimer were employed in this work. Using the second model, K_a between host and guest can be determined with a known host dimerisation constant (K_{dim}) by measuring the change in NMR chemical shift of a probe on the host in a set of solutions with constant host concentration and increasing guest concentration.[155]

The system equilibrium, and the corresponding binding constant equations are:

$$H + G \Leftrightarrow HG \qquad K_a = [HG] / [H][G] \qquad (1)$$
$$H + H \Leftrightarrow H_2 \qquad K_{dim} = [H_2] / [H]^2 \qquad (2)$$

The mass balances for the system are:

$$[H]_0 = [H] + [HG] + 2[H_2] \qquad (3)$$
$$[G]_0 = [G] + [HG] \qquad (4)$$

From equations (1) – (3), expressions can be derived for the concentration of all species in solution, for each titration point:

$$[H_2] = K_{dim}[H]^2 \qquad (5)$$
$$[HG] = [H]_0 - [H] - 2[H_2] \qquad (6)$$
$$[G] = [HG] / [H]\, K_a \qquad (7)$$

The NMR probe on the Host has 3 chemical shifts corresponding to the species in solution: the unbound chemical shift, δ_H, the chemical shift of the complex with the guest, δ_{HG}, and the chemical shift of any Host dimer that forms, δ_{H2}. The complex and the dimer chemical shifts are assumed to be the same, $\delta_{HG} = \delta_H$. The observed chemical shift will be a mixture of the 3 shifts according the mole fraction of each species present and can be calculated according to:

$$\delta_{calc} = (\delta_H[H] + \delta_{HG}[HG] + 2\delta_{H2}[H_2]) / ([H] + [HG] + 2[H_2]) \qquad (8)$$

The program searches for values of K_a , δ_H, and δ_{H2} (= δ_{HG}) that give δ_{calc} values that most closely match the experimental chemical shift, δ_{obs}, for each point in the titration curve. This process is described below.

The chemical shift of the Host dimer, δ_{H2}, and its binding constant, K_{dim}, is determined previously in an independent experiment. The program guesses a value for K_a, δ_H, and δ_{H2} (= δ_{HG}) then, for each point in the titration:

i. Use a numerical approach (Newton-Raphson) to find the value of free Host, [H] where:

$[G]_0 - [G] - [HG] = 0$, where $[G]_0$ is known, and [G] and [HG] are given by (6) and (7).

ii. Use (5) to calculate the concentration of $[H_2]$.

iii. Use (6) to calculate the concentration of [HG].

iv. Use (8) to calculate the expected chemical shift for this point in the titration.

v. If it is not a good match try new values of K_a, δ_H, and δ_{H2} (= δ_{HG}).

Typically one or two different proton resonances were monitored at the same time giving the corresponding data sets. The association constant for a single run was calculated as the mean of the values obtained for each of the signals followed during the titration, weighted by the observed changes in chemical shift. The association constants from different runs were then averaged.

b) Thordarson Global Fitting (^{1}H NMR).

In some cases, when there were two shifting NH/NH_2 nuclei, the 1:1 binding constants were obtained from the ^{1}H NMR titration experiments using a custom written global nonlinear regression analysis program developed by P. Thordarson within the Matlab R2012b package utilizing the Simplex algorithm.[157] This fitting method uses a global approach that considers both set of data simultaneously, which enhances the quality of the fitting procedure. However, host (nor guest) dimerization were considered using this fitting approach.

The standard errors (SEy) are calculated by:

$$\mathrm{SE}_y = \sqrt{\frac{\sum(y_{data} - y_{calc})^2}{N - k}}$$

Where *N* is the number of data points and *k* the number of parameters to be fitted.

c) *ReactLab EQUILIBRIA* (UV-vis).

ReactLab EQUILIBRIA, a more sophisticated version of the previous SpecFit program, is a program developed and commercialized by Jplus Consulting Pty Ltd (http://jplusconsulting.com/ ; 8 Windsor Road, East Fremantle, WA 6158, Australia). It allows for the global fitting of multi-wavelength spectroscopic data in equilibrium titration measurements to chemical reaction schemes, and determines all equilibrium constants in the underlying mechanism. *ReactLab*™ algorithms fit complete reaction models directly to multivariate data and delivers all the required parameters in one step. The analysis also yields the concentration distributions of all species and the individual spectra of all the participating species. The program, including all algorithms and the GUI frontend has been developed in Matlab and compiled to produce the final deployable application.

A large wavelength region in the absorption spectra (from 250 to 450 nm; each wavelength representing one set of data) was fitted by this software. However, only a few selected wavelengths are plotted in the charts below. Both host and guest dimerization constants were considered in the analysis of the 1:1 host-guest binding constants.

[157] a) J. A. Nelder, R. Mead, *Comp. J.* **1965**, *7*, 308–313; b) J. C. Lagarias, J. A. Reeds, M. H. Wright and P. E. Wright, *SIAM J. Optim.* **1998**, *9*, 112–147.

Chapter 3

Cyclic Tetramer Self-Assembly in solution

The goal of this Chapter is the study of the Ring-Chain equilibria (see Introduction) of our target monomers. For that purpose, the simplest systems were considered first: ditopic monomers carrying complementary bases at the edges (Figure 3.1, **G**:**C**, **iG**:**iC**, **A**:**U**) so that they form cyclic tetramers composed of one component linked through the simplest oligophenylene-ethynylene central block (**B1**). Since we can play with three different pairs of nucleobases, we aim to study which is the role of the symmetry of multipoint hydrogen bonding on chelate cooperativity of cyclization processes because of it has never been addressed before.

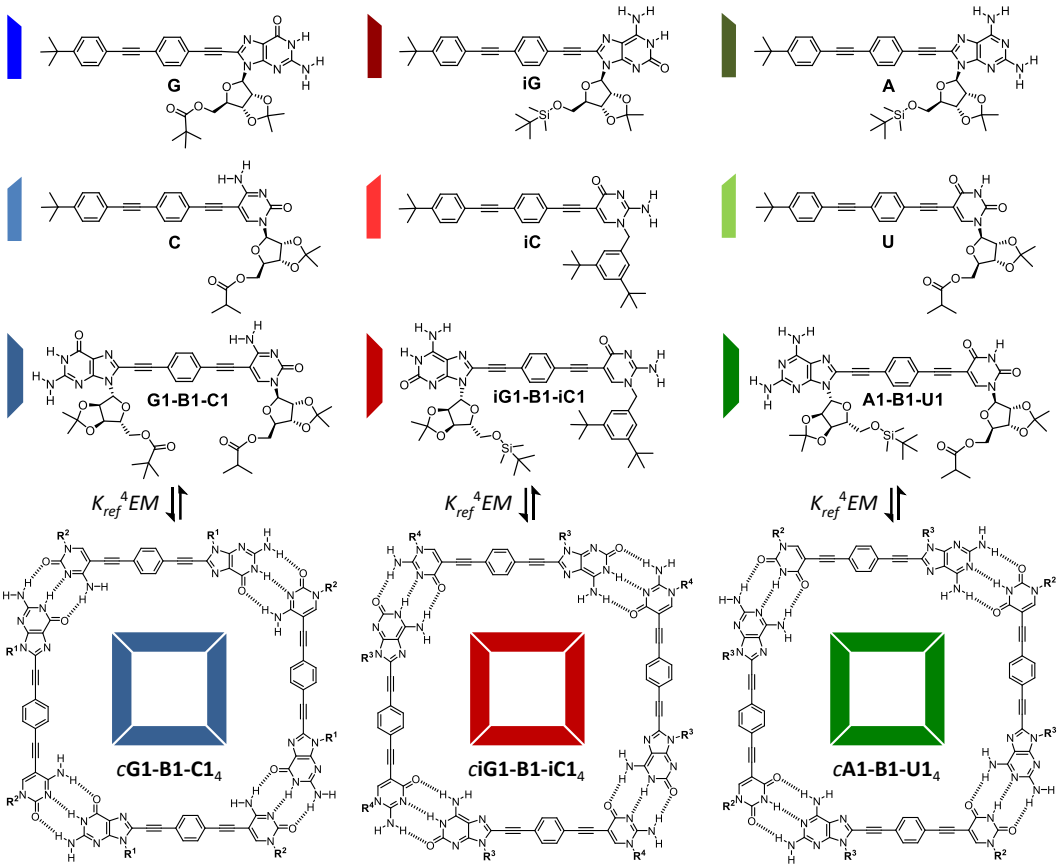


Figure 3.1. Structure of lipophilic dinucleoside **G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-C1**, reference mononucleoside compounds, **G**, **C**, **iG**, **iC**, **A** and **U**, and the three cyclic tetramers formed in solution *c***G1-B1-C1**$_4$ (*ADD-DAA*), *c***iG1-B1-iC1**$_4$ (*DDA-DDA*) and *c***A1-B1-C1**$_4$ (*DAD-ADA*).

In this chapter, we have devised several experiments that demonstrate the consequences of a suitable monomer design (Figure 3.1) in the fidelity of the self-assembly process and in the thermodynamic and kinetic stability of ring-closed structures (*c***M**$_4$) when compared to linear assemblies. The self-assembly of monomers **G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-U1** was analyzed in different solvents by a wide number of concentration- and temperature-dependent spectroscopic methods (1D and 2D ^{1}H NMR, as well as absorption, emission and CD spectroscopy). All of these data can be found as well in the Supporting

information of our papers. *Angew. Chem. Int. Ed.* **2015**, *54*, 6780–6784 (Ref: 156) and *Angew. Chem. Int. Ed.* **2016**, *55*, 223–227 (Ref:158)

All the optical spectroscopy measurements were carried out by Jorge Camacho-García and form part of his Thesis, which has been developed at the same time that this one. We thought appropriate include these studies here in order to assess completely the stability of our systems in organic solvents at high dilutions.

3.1 Results and Discussion.

3.1.1 Synthetic Strategy to Unsymmetric Complementary Dinucleosides.

In Section **1.1.**, we have commented the design and synthesis of main components of the tetrameric species. Here, will study the simplest cyclic system formed by only one component driven by the association of self-complementary nucleobases at both ends. For that purpose, the monomers discussed in this section have been prepared through two consecutive Sonogashira reactions from the molecular fragments prepared previously. As it is shown in Figure 3.2, "base 1" is most frequently a pyrimidine and the purine is incorporated later as "base 2", since its manipulation is more delicate for solubility reasons and it is typically more valuable. Since, in the first step the monocoupling product is targeted, more equivalents of the central block than the ethynylated nucleobase derivative are used in order to obtain the desired product with the maximum yield. Through the use of this strategy, the formation of small amounts of the di-substituted product as well as the homocoupling product have been solved.

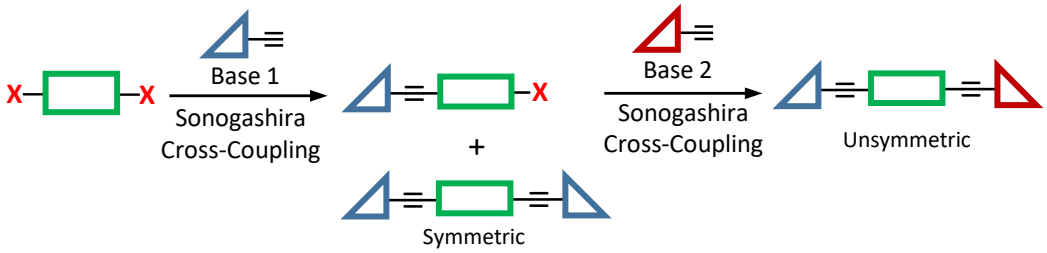


Figure 3.2. General synthetic pathway toward final target monomers.

3.1.2 Study of the Self-Assembly by ^{1}H NMR.

Cyclic Tetramer Association in Apolar Solvents.

A first insight into the self-assembly of monomers **G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-U1** was obtained from the analysis of ^{1}H NMR data in the most common apolar solvents and then compared it with the respective association process from free nucleobases in solution. For instance, Figure 3.3 displays the different stabilities between *c***G1-B1-C1**$_4$ and the **G1**+**C1** 1:1 complex in different solvents. The characteristic signals for **G**-**C** H-bonding are the **G1** amide and **C1** amine that are found around 13.5 and between 10.0-9.0 ppm (this last proton signal is more exposed and thus more sensitive to the solvent environment) respectively. As it is evidenced *c***G1-B1-C1**$_4$ displays the same ^{1}H NMR spectra in all the solvents suggesting the formation of a single species in solution. In contrast, solvents like DMF-d_7 and DMSO-d_6 that

[158] C. Montoro-García, J. Camacho-García, A. M. López-Pérez, M. J. Mayoral, N. Bilbao, D. González-Rodríguez, *Angew. Chem. Int. Ed.* **2016**, *55*, 223–227.

compete strongly for H-bonding can dissociate the tetramer as it is shown in Figure 3.3b revealing the correspondent **G1** amide bonded signal to these solvents. On the other hand, Figure 3.3a shows the different stability of **G1**+**C1** systems. Here, polar solvents like THF-d_8 or DMF-d_7 upfield these characteristic signals (**G1** amide and **C1** amine) for the **G**+**C** complex due to they are less involved in intermolecular H-bonding than the *c***G1-B1-C1**$_4$ tetramer. Here, is interesting to note the thermodynamic and kinetic differences between these associations. **G1**+**C1** system reveals a thermodynamic stability in chlorinated solvents due to the presence of the same spectra suggest that the dimer is the only one structure in solution. In contrast, when more polar solvents were employed, the system showed an upfield **G1** amide signal that is attributed to a fast-exchange equilibria in the ^{1}H NMR timescale between dimers and monomers. On the other hand, **G1-B1-C1** exhibits an impressive thermodynamic stability in all the solvents except in DMSO-d_6, where the signals attributed to **G**-**C** H-bonding do not change in shape or chemical shift. Furthermore, in DMF-d_7 **G1-B1-C1** reveals two set of signals corresponding to the monomer-tetramer equilibria in slow-exchange in the ^{1}H NMR timescale, suggesting a very slow kinetic behavior, which will be discussed later.

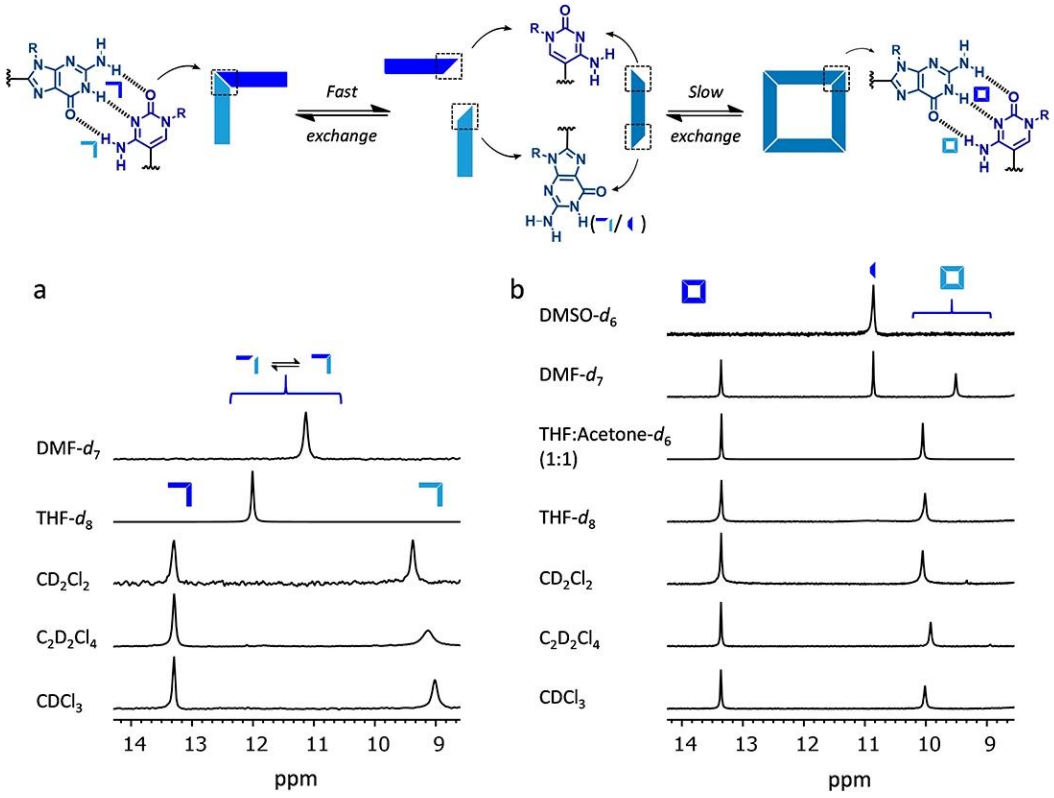


Figure 3.3. *Self-assembly in different solvents.* Region of the ^{1}H NMR spectra of (**a**) **G1-B1-C1** monomer and (**b**) **G**+**C** 1:1 complex in different solvents. In all cases C = 1.0 x 10^{-2} M, T = 298 K.

10^{-1}-10^{-5} M solutions of **G1-B1-C1** and **iG1-B1-iC1** in these solvents displayed ^{1}H NMR spectra (Figure 3.4b) that are characteristic of base pair association. The H-bonded **iG** amide and **iC** amine proton signals appeared in $CDCl_3$ at 13.7 ppm and 10.1 ppm for **iG1-B1-iC1** that are very similar to **G1-B1-C1** as was commented previously. For these two monomers, the shape and position of these H-bonded signals, in particular the more shielded G or iG amide protons, do not change significantly with concentration (Figure 3.4b). This behaviour contrasts

the changes observed in 1:1 mixtures of mononucleosides **G** and **C** or **iG** and **iC**, where the H-bonded protons experience an upfield shift upon increasing dilution as the paired bases dissociate (Figure 3.4a). The exceptional stabilization of the H-bonded species in **G1-B1-C1** and **iG1-B1-iC1** was attributed to the quantitative formation of the cyclic tetramers *c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$ in apolar solvents.

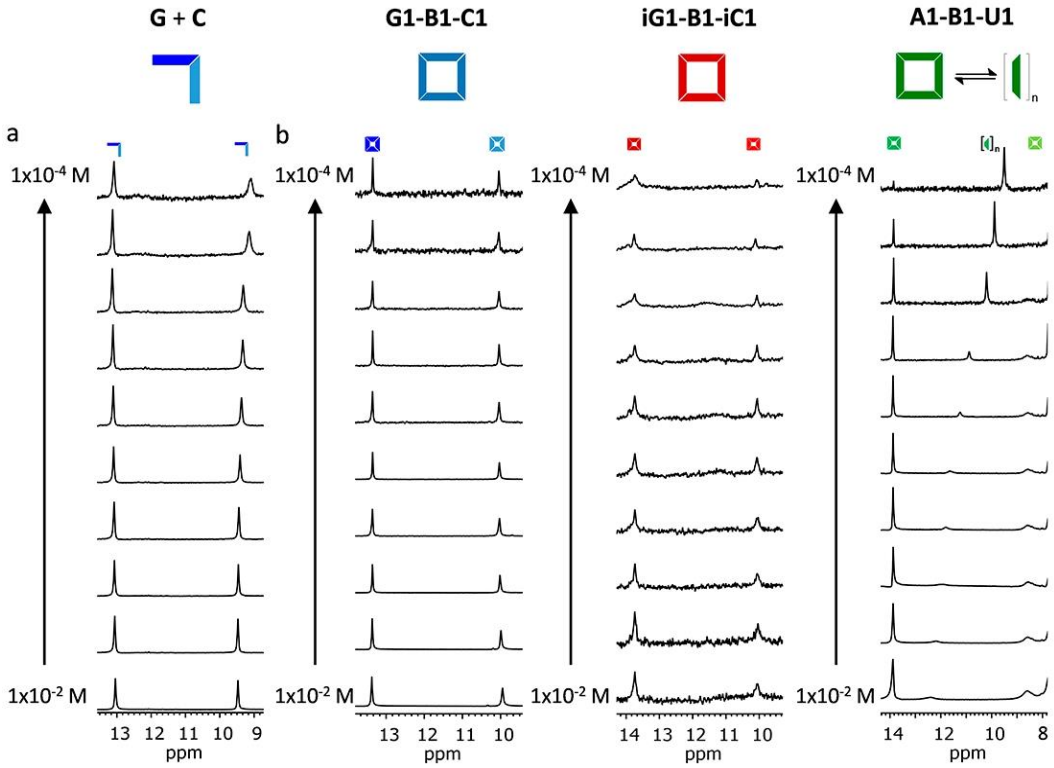


Figure 3.4. *Concentration-dependent measures.* 14-8 ppm region of the ^{1}H NMR spectra of (**a**) **G**+**C** 1:1 complex and (**b**) *c***G1-B1-C1**$_4$, *c***iG1-B1-iC1**$_4$ and *c***A1-B1-U1**$_4$ in $CDCl_3$ at different concentrations. *T* = 298 K in all cases.

Temperature-dependent measurements (Figure 3.5b) revealed that only in the case of the **G**+**C** complex, the H-bonded G amide and C amine proton signals shift upfield at higher temperatures, as the complex dissociates, but not those of the *c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$ cyclic tetramers which kept in all the temperatures the same distribution. G and iG amine protons are typically observed at temperatures below 278 K at 8.5 and 10.0 ppm, respectively due to a more restricted rotation around the C-N bound.[134]

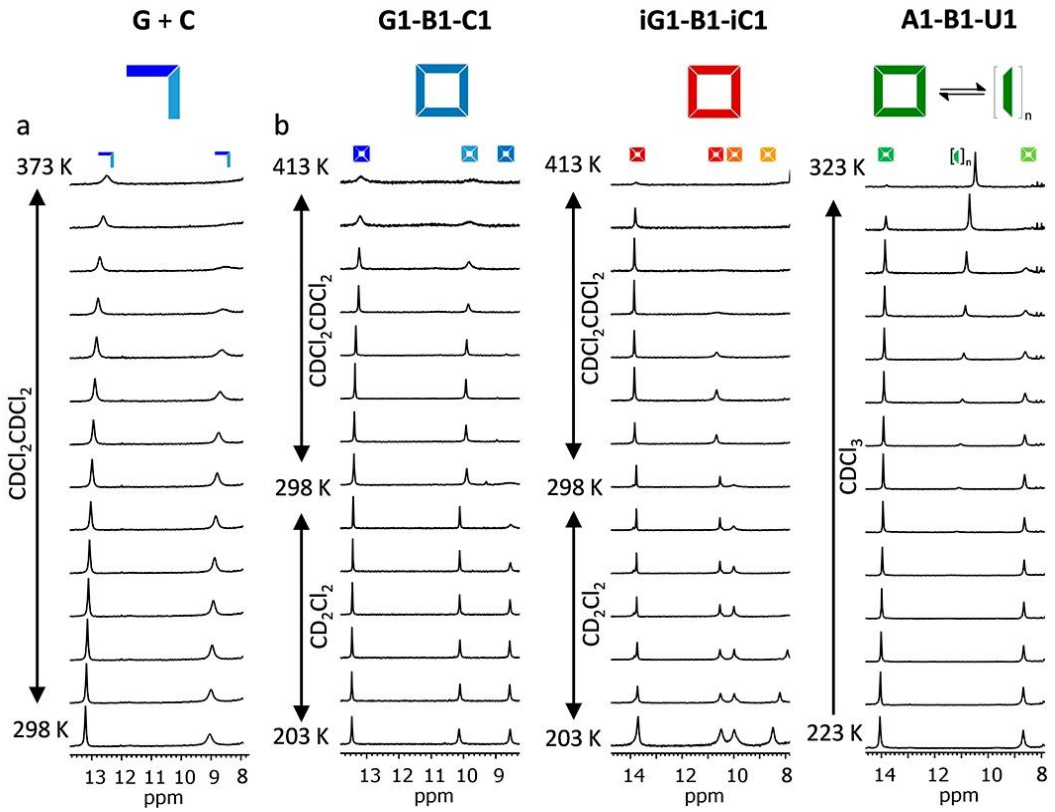


Figure 3.5. *Temperature-dependent measures*. 14-8 ppm region of the ^{1}H NMR spectra of (**a**) **G**+**C** 1:1 complex and (**b**) *c***G1-B1-C1**$_4$, *c***iG1-B1-iC1**$_4$ and *c***A1-B1-U1**$_4$ in $CDCl_3$ at different temperatures. In all cases *C* = 1 x 10^{-2} M.

Base pairing in *c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$ was confirmed through NOESY experiments (Figure 3.6a,b), where clear cross-peaks between G-H^1 and C-H^1, and between iG and iC, were observed.

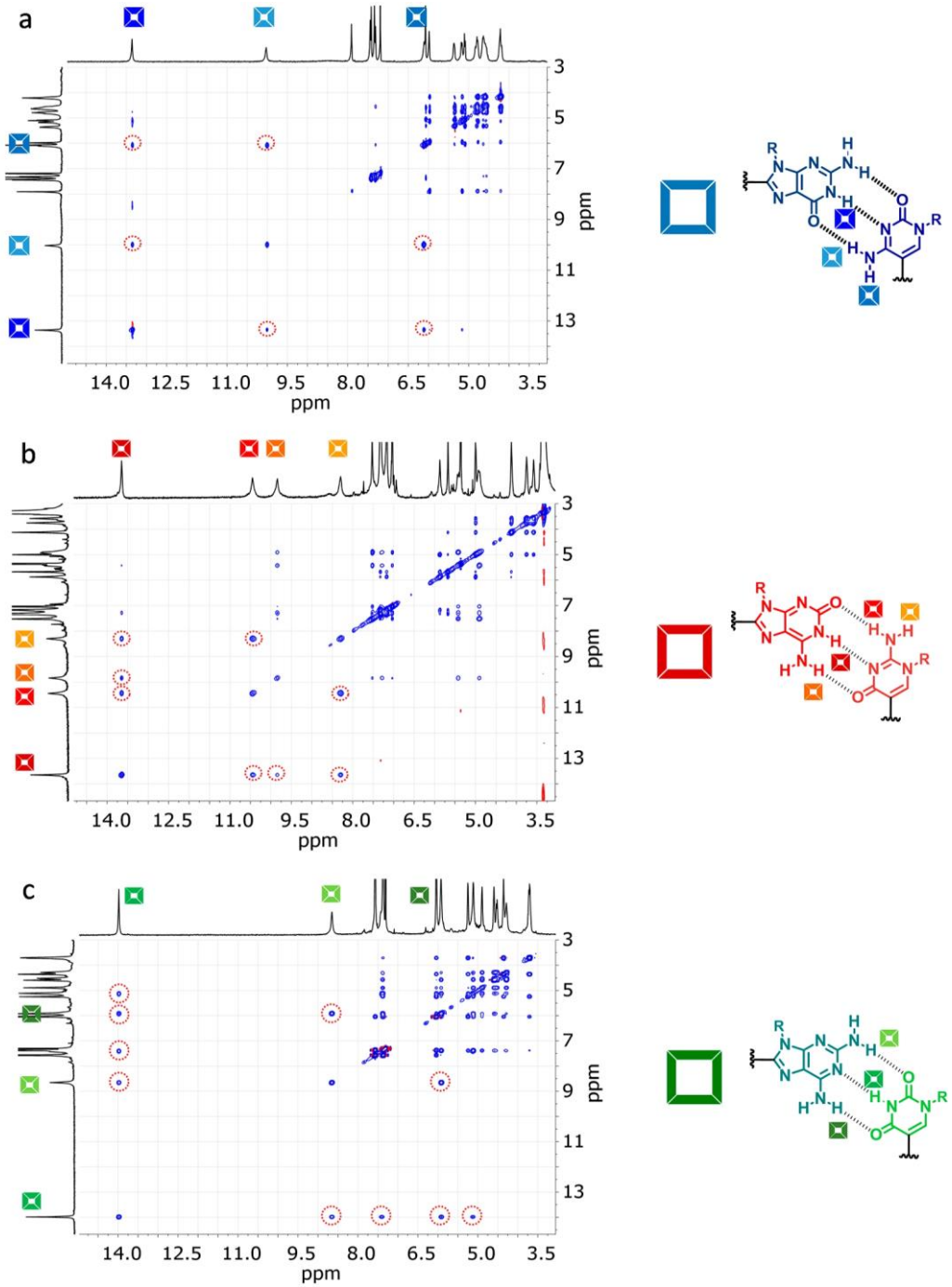


Figure 3.6. *Base-pairing confirmation by 2D NOESY spectra*. Region of the NOESY NMR spectrum of (**a**) *c***G1-B1-C1**$_4$ (*C* = 1.0 x 10^{-2} M and *T* = 298 K), (**b**) *c***iG1-B1-iC1**$_4$ (*C* = 1.0 x 10^{-2} M and *T* = 238 K) and (**c**) *c***A1-B1-U1**$_4$ in $CDCl_3$ (*C* = 1.0 x 10^{-2} M and *T* = 253 K), showing cross-peaks between the H-bonded proton signals.

DOSY experiments carried out in $CDCl_3$ (Figure 3.7a), on the other hand, revealed in both cases a single diffusing species. The diffusion coefficients obtained from the T_1/T_2 relaxation curves match an assembly whose hydrodynamic radius agrees with the expected

size of a cyclic tetramer (see Table 3.1), obtained from molecular modeling studies (Figure 3.7c). A spherical model was employed in the calculation of the radii of the tetramer due to it fits properly with the experimental data and has been used previously with flat self-aseembly species.[110] The model structures also show that only a cyclic tetramer allows for optimal triple H-bonding interactions between the nucleobase edges. Cyclic trimer or pentamer assemblies would be more strained and far from achieving an optimal Watson-Crick H-bonding geometry.

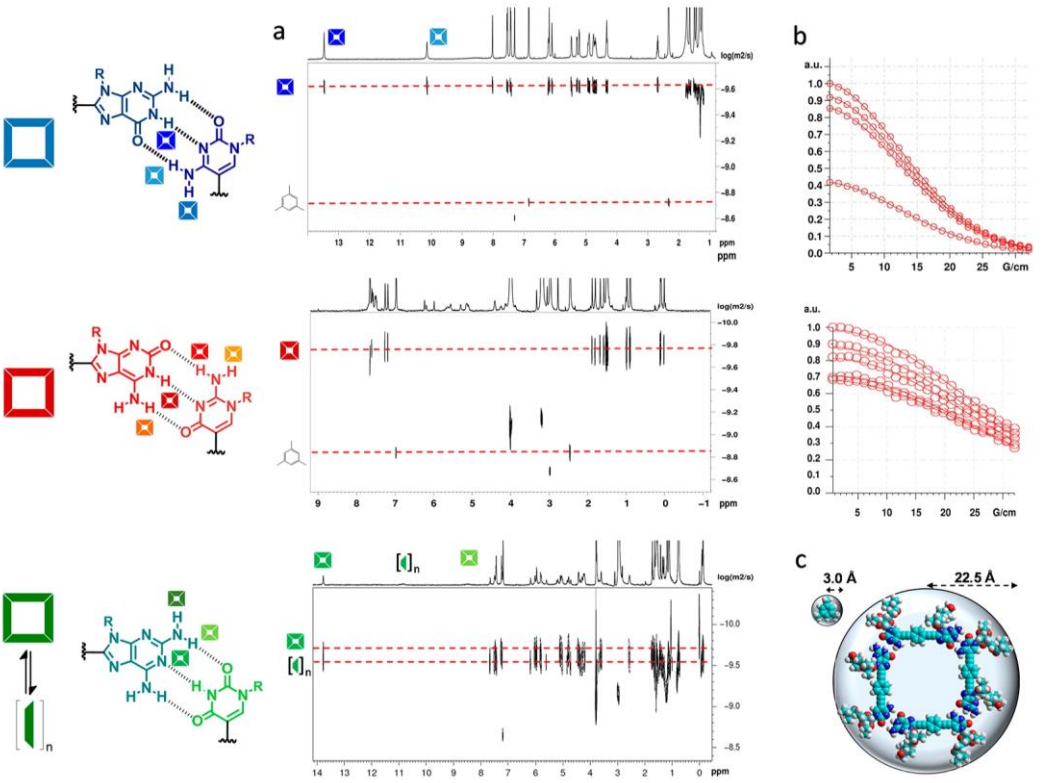


Figure 3.7 *Size estimation through 2D-DOSY NMR spectra* (**a**) 2D-DOSY NMR spectrum of *c***G1-B1-C1**$_4$, *c***iG1-B1-iC1**$_4$ and *c***A1-B1-U1**$_4$ in $CDCl_3$ (C = 1.0 x 10^{-2} M, T = 298 K, in all cases) with mesitylene as internal reference. (**b**) T_1/T_2 relaxation curves of the ^{1}H NMR signals from which D_0 was calculated. *c***G1-B1-C1**$_4$ at 8.02, 7.54, 7.44 and 5.47 ppm. *c***iG1-B1-iC1**$_4$ at 5.59, 6.52, 1.22, and 1.12 ppm (**c**) Molecular radii estimated from computed models of mesitylene and cyclic tetramer *c***G1-B1-C1**$_4$. The structures were optimized by PM3 semiempirical calculations using the HyperChem 8.0.3 software package.

The experimental values of diffusion coefficient (D_0) obtained from the DOSY NMR experiments (Figure 3.7b) can be related to the hydrodynamic radius (R) of a molecule with average spherical size through the Stokes-Einstein equation:[159]

$$D = \frac{k_b T}{6\pi\eta_0 R}$$

where k_b is the Boltzmann's constant, T the temperature and η_0 is the viscosity of the solvent. The use of an internal reference whose hydrodynamic radius is known, like mesitylene

[159] A. Einstein, *Ann. Phys.* **1906**, *324*, 289–306.

in this case (R_{ref} = 3.0 Å), allows the estimation of the aggregate size by direct D_0 comparison, using the following equation:[160]

$$R_o = \frac{R_{ref}\, D_{ref}}{D_0}$$

M	Peak ppm	D_0 m^2s^{-1}	R[a] Å	R[b,c] Å
***c*G1-B1-C1$_4$[c]**	8.02	8.632x10^{-12}	**22.46**	22.5
	7.54	8.608x10^{-12}	**22.53**	
	7.44	8.587x10^{-12}	**22.58**	
	5.47	8.459x10^{-12}	**22.92**	
***c*iG1-B1-iC1$_4$**[c]	6.59	2.311x10^{-12}	**25.84**	
	6.52	2.260x10^{-12}	**26.42**	
	1.22	2.390x10^{-12}	**24.98**	
	1.12	2.470x10^{-12}	**24.18**	
	0.99	2.387x10^{-12}	**25.02**	
***c*A1-B1-U1$_4$**[d]	7.62	6.817x10^{-12}	**23.61**	
	5.17	6.640x10^{-12}	**24.24**	
	4.33	6.589x10^{-12}	**24.43**	
	3.70	6.725x10^{-12}	**23.93**	
Mesitylene[c]	6.98	1.968x10^{-11}		3.0
	2.47	2.013x10^{-11}		
Mesitylene[d]	6.80	5.400x10^{-11}		
	2.32	5.330x10^{-11}		

[a] Calculated using mesitylene as an internal reference. [b] Calculated from computed models. [c] in $CDCl_3$. [d] in 2:3 *v/v* $CDCl_3$-CCl_4 (see Figure 3.9c).

Table 3.1. Calculated diffusion coefficients (D_0) and estimated hydrodynamic radii (R).

Finally, ESI Q-TOF mass spectrometry experiments (positive mode) also sustained the formation of ***c*G1-B1-C1$_4$** in a 1:2 *v/v* $CHCl_3$-MeCN mixture, and we could detect the singly, doubly and triply-charged ***c*G1-B1-C1$_4$** peaks and some of its fragments as it is shown in Figure 3.8.

[160] a) P. Timmerman, J.-L. Weidmann, K. A. Jolliffe, L. J. Prins, D. N. Reinhoudt, S. Shinkai, L. Frish, Y. Cohen, *J. Chem. Soc*., *Perkin Trans. 2*, **2000**, 2077–2089; b) Y. Cohen, L. Avram, L. Frish, *Angew. Chem. Int. Ed.* **2005**, *44*, 520–554, *Angew. Chem.* **2005**, *4*, 524–560.

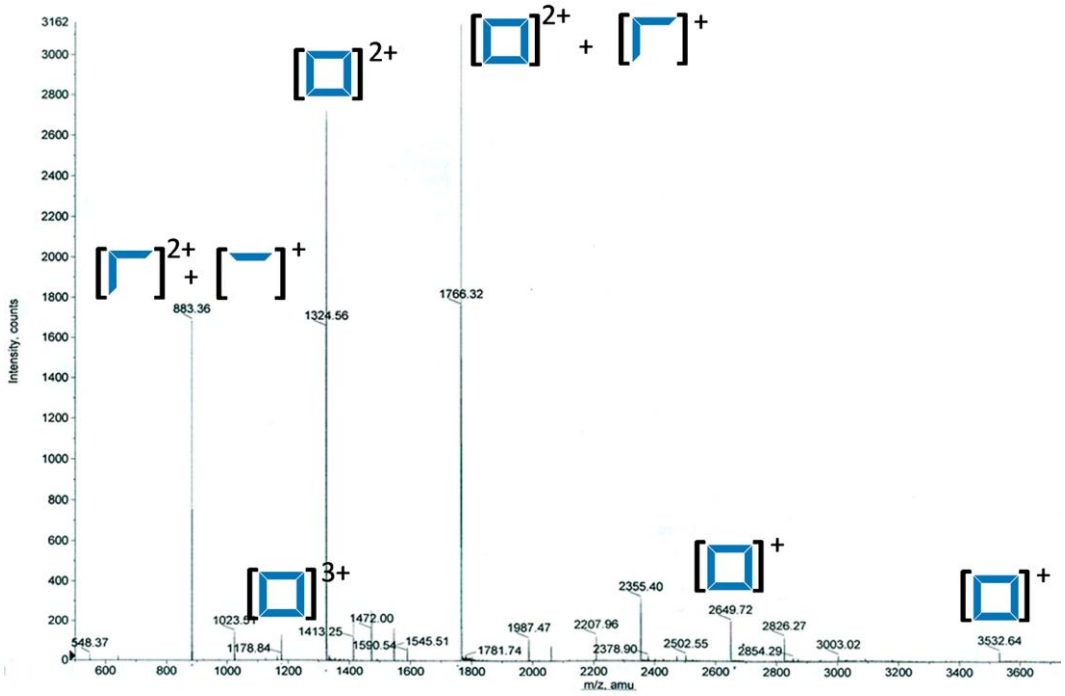


Figure 3.8. ESI Q-TOF mass spectrometry (positive mode) of **G1-B1-C1** in a 1:2 *v/v* $CHCl_3$-MeCN mixture.

The Special Case of A1-B1-U1.

The association behavior of monomer **A1-B1-U1** in the same conditions is markedly different. The ^{1}H NMR spectra in $CDCl_3$ solutions revealed a mixture of two main species (Please, come back to Figure 3.4b and Figure 3.5b). One of them exhibits a U-H^1 imide proton signal at 13.8 ppm, whose shape and position does not change significantly with concentration or temperature. The other species presents a U-H^1 signal within the 13.0-9.0 ppm range and exhibits a high sensitivity to concentration or temperature changes. This picture was interpreted as a slow equilibrium between *c***A1-B1-U1**$_4$ and monomer **A1-B1-U1**, which is at the same time in fast equilibrium with small open oligomers (dimers, trimers,..). The first unshifted U-H^1 signal at 13.8 ppm was therefore assigned to *c***A1-B1-U1**$_4$ and the second imide signal to the mixture of small, rapidly exchanging oligomers, whose behavior upon concentration or temperature changes is reminiscent to the one shown by the 1:1 complex of **A** and **U** in the same conditions. DOSY experiments (Figure 3.7a) in $CDCl_3$ at 298 K also suggested the presence of two main species in slow equilibrium that displayed distinct diffusion coefficients.

Decreasing concentration (Figure 3.4b) or increasing temperature (Figure 3.5b) results in the growth and upfield shift of the U-H^1 signal between 13.0-9.0 ppm at the expense of the *c***A1-B1-U1**$_4$ signal at 13.8 ppm, as the **A**-**U** base pairs are less involved in H-bonding and the **A1-B1-U1** monomer gradually becomes the most abundant species in solution. At the other extreme, at high concentrations and low temperatures, the ^{1}H signal at 13.8 grows to become the only detectable imide signal, indicating that **A1-B1-U1** is mainly associated as a cyclic tetramer. The analysis of these samples by NOESY (Figure 3.6c) revealed cross-peaks between U-H^1 and A-H^2 proton signals, confirming H-bonding association.

The stabilization of the *c***A1-B1-U1**$_4$ macrocycle could also be achieved by decreasing solvent polarity in mixtures of $CDCl_3$ and CCl_4 (Figure 3.9). As shown in Figure 3.9a, the increase in the volume fraction of CCl_4 produced the gradual disappearance of the **A1-B1-U1** monomer signals and the quantitative formation of *c***A1-B1-U1**$_4$, which revealed now a single set of ^{1}H NMR signals. NOESY (Figure 3.9b) and DOSY (Figure 3.9c) experiments performed in a 2:3 *v/v* $CDCl_3$-CCl_4 mixture at 298 K were now very similar to those obtained for *c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$ in $CDCl_3$, and supported the presence of a predominant H-bonded species, diffusing with a size that matches the predicted *c***A1-B1-U1**$_4$ hydrodynamic radius obtained from computational models (see Figure 3.9c,d and Table 3.1).

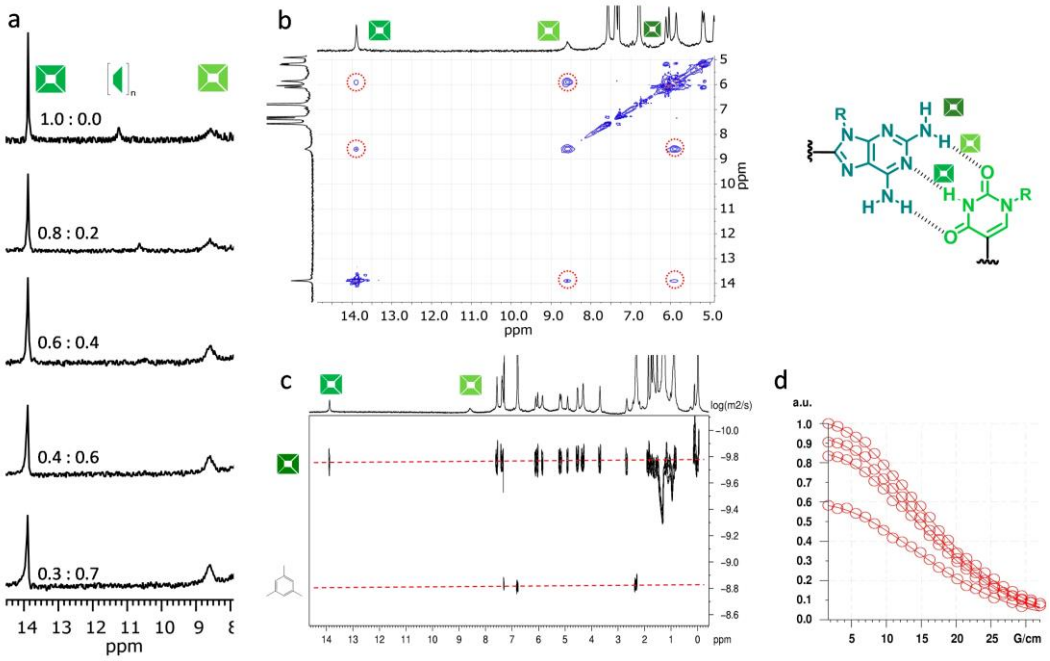


Figure 3.9. *Stabilization of the c**A1-B1-U1**$_4$ macrocycle.* (**a**) Changes in the ^{1}H NMR spectra of **A1-B1-U1** upon increasing the volume fraction of CCl_4 in $CDCl_3$. (**b**) NOESY and (**c**) 2D-DOSY NMR spectra of **A1-B1-U1** with relaxation T_1/T_2 curves in 2:3 *v/v* $CDCl_3$-CCl_4. In all cases, C = 1.0 x 10^{-2} M, T = 298 K.

Dilution experiments in this 2:3 *v/v* $CDCl_3$-CCl_4 mixture (Figure 3.10) revealed the growing of the A1-B1-U1 monomer signals only at concentrations well below 10^{-3} M.

On the other hand, the addition of polar cosolvents produced the dissociation of the cyclic tetramers and allowed us to study in further detail the monomer–tetramer equilibria. Due to the different association strength of the *c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$ assemblies, on one hand, and the *c***A1-B1-U1**$_4$ species, on the other, we employed quite different conditions in each case.

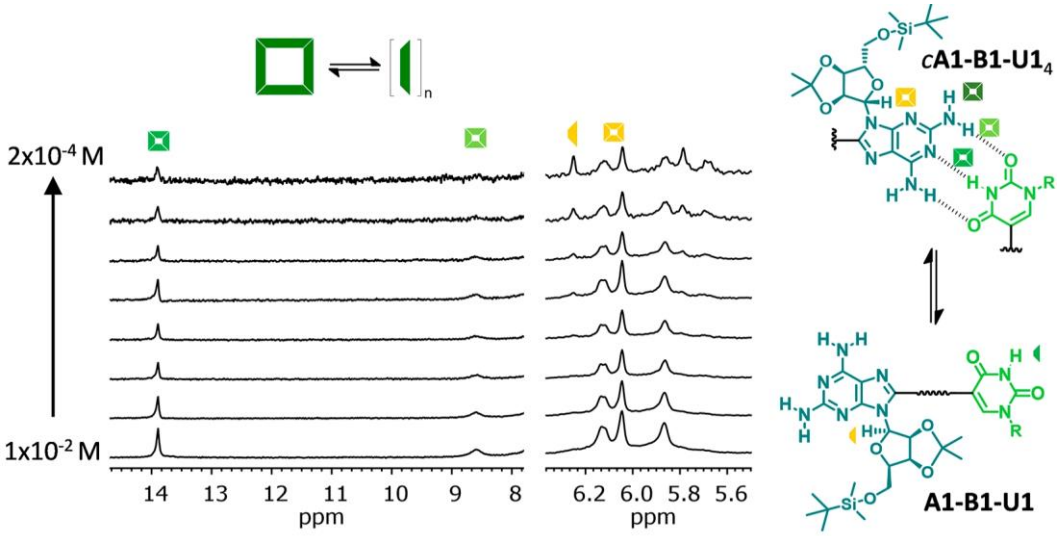


Figure 3.10. *Concentration-dependent measures.* 14.0-8.0 ppm and 6.3-5.6 ppm regions of the ^{1}H NMR spectra of **A1-B1-U1** upon decreasing the concentration of **A1-B1-U1** in 2:3 *v/v* $CDCl_3$-CCl_4, *T* = 298 K.

Addition of Polar Cosolvents. Monomer-Tetramer Equilibria (M_4-*c*M_4).

Firstly, we monitored the changes observed by ^{1}H NMR upon increasing DMSO-d_6 content in $CDCl_3$ solutions (Figure 3.11). *c***iG1-B1-iC1**$_4$ shows also a high resistance to this highly polar cosolvent like *c***G1-B1-C1**$_4$, and the cyclic tetramer persisted even after the addition of *ca.* 80% DMSO-d_6. The *c***A1-B1-U1**$_4$ assembly, on the contrary, could not resist more than 7% DMSO in $CDCl_3$ and 12% DMSO in the 2:3 *v/v* $CDCl_3$-CCl_4 mixture. In this case, acetone-d_6 was instead employed (Figure 3.11b) as the polar cosolvent. It is interesting to note that the addition of small amounts of this solvent onto $CDCl_3$ solutions affected mainly the position of the U-H^1 signal assigned to the **A1-B1-U1** monomer and small oligomers mixture in fast equilibrium. Further acetone-d_6 addition (over *ca.* 60%) eventually led to the complete dissociation of the *c***A1-B1-U1**$_4$ assembly (Figure 3.11c), monitored by the disappearance of the signal at 13.8 ppm. Throughout all these titrations with polar solvents the tetramer proton signals (*c***G1-B1-C1**$_4$, *c***iG1-B1-iC1**$_4$ or *c***A1-B1-U1**$_4$) did not change significantly in shape or position, suggesting in all cases slow equilibria in the NMR timescale and underlining the superior kinetic stability of this cyclic species.

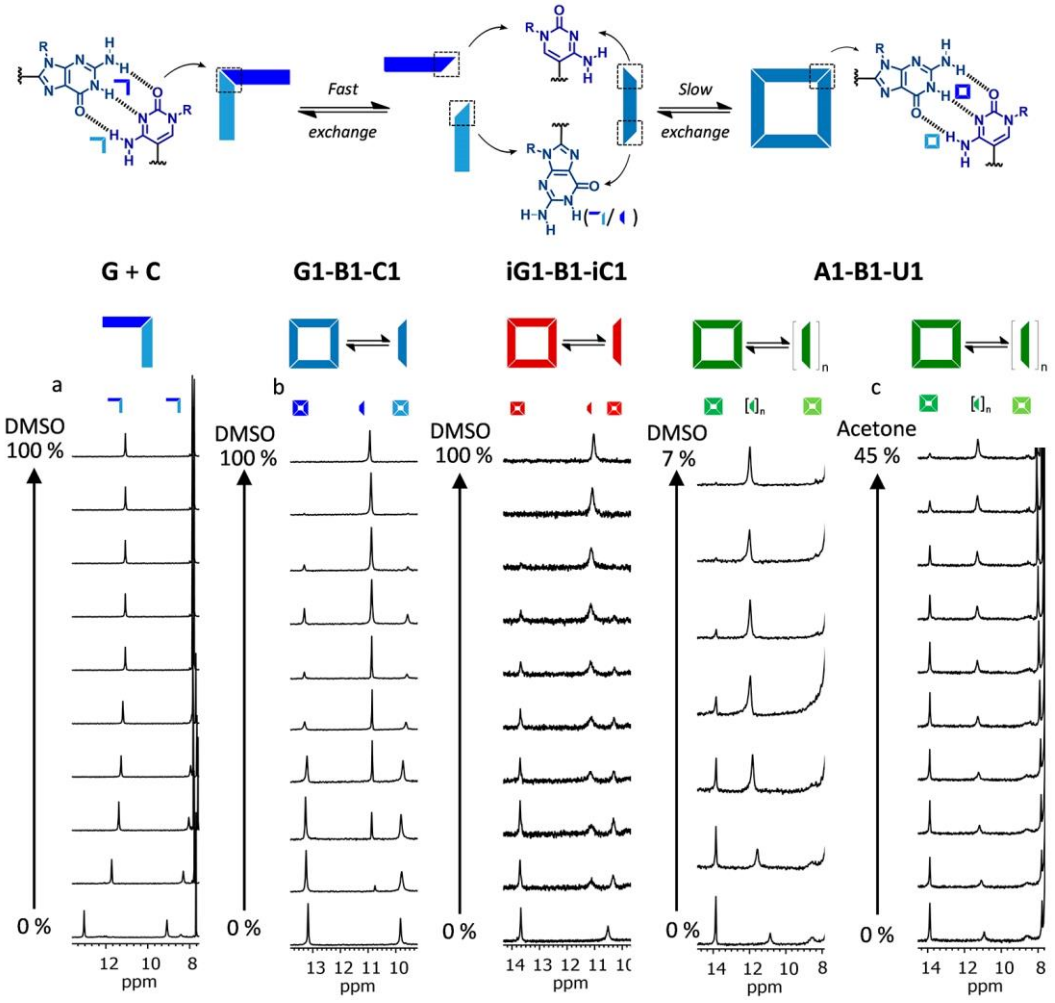


Figure 3.11. *Increasing amounts of DMSO content.* 14-8 ppm region of the ^{1}H NMR spectra of (**a**) **G**+**C** 1:1 complex and (**b**) **G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-U1** upon increasing the volume fraction of DMSO-d_6 in $CDCl_3$, (**c**) **A1-B1-U1** upon increasing the volume fraction of acetone-d_6 in $CDCl_3$ (C = 1 x 10^{-2} M, T = 298 K in all cases).

*c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$ self-assembly was also studied in 100% DMF-d_7[161] solutions within the 10^{-1}-10^{-3} M range, where a slow monomer-tetramer equilibrium was again noted (Figure 3.12). The **A1-B1-U1** dinucleoside, on the contrary, existed only as monomer in these conditions. A mixture of 5:1 *v/v* $CDCl_3$-acetone-d_6 was used in order to reduce the population of non-cyclic H-bonded oligomers and enhance the all-or-non behavior. On the other hand, concentration-dependent measurements in the solvents previously commented revealed the presence of an equilibrium between monomer and cyclic tetramer. It is interesting to note that the shape and position of the **G** (or **U**)-amide and **C** (or **A**)-amine do not change with concentration, suggesting a very low exchange in the NMR timescale and an "all or nothing" behavior. The concentrations of the respective species in solution *c***G1-B1-**

[161] Both monomers revealed the same behavior in concentration- and temperature-dependent measures in a 1:1 *v/v* $CDCl_3$-DMSO-d_6 mixture but finally we discarded this option to carry out all the pertinent measurements because of chloroform is a highly volatile solvent and the stability of our systems are highly influenced by the ratio *v/v* $CDCl_3$-DMSO-d_6.

C1$_4$**/G1-B1-C1**, *c***iG1-B1-iC1**$_4$**/ iG1-B1-iC1** and *c***A1-B1-U1**$_4$ **/A1-B1-U1** were calculated in each spectrum by signal integration (at least 2 C-H proton signals for each species were averaged). Within the whole concentration range, the process follow the linear relationship $[M_4] = K_T[M]^4$ (but not for those formed by a more strained trimer or pentamer species, supporting the formation of the tetramer as the most stable structure), which allowed to extract the equilibrium constants (K_T; Table 3.3, in page 133):

$$K_T = \frac{[c\mathbf{G1B1C1_4}]}{[\mathbf{G1B1C1}]^4} \qquad K_T = \frac{[c\mathbf{iG1B1iC1_4}]}{[\mathbf{iG1B1iC1}]^4} \qquad K_T = \frac{[c\mathbf{A1B1U1_4}]}{[\mathbf{A1B1U1}]^4}$$

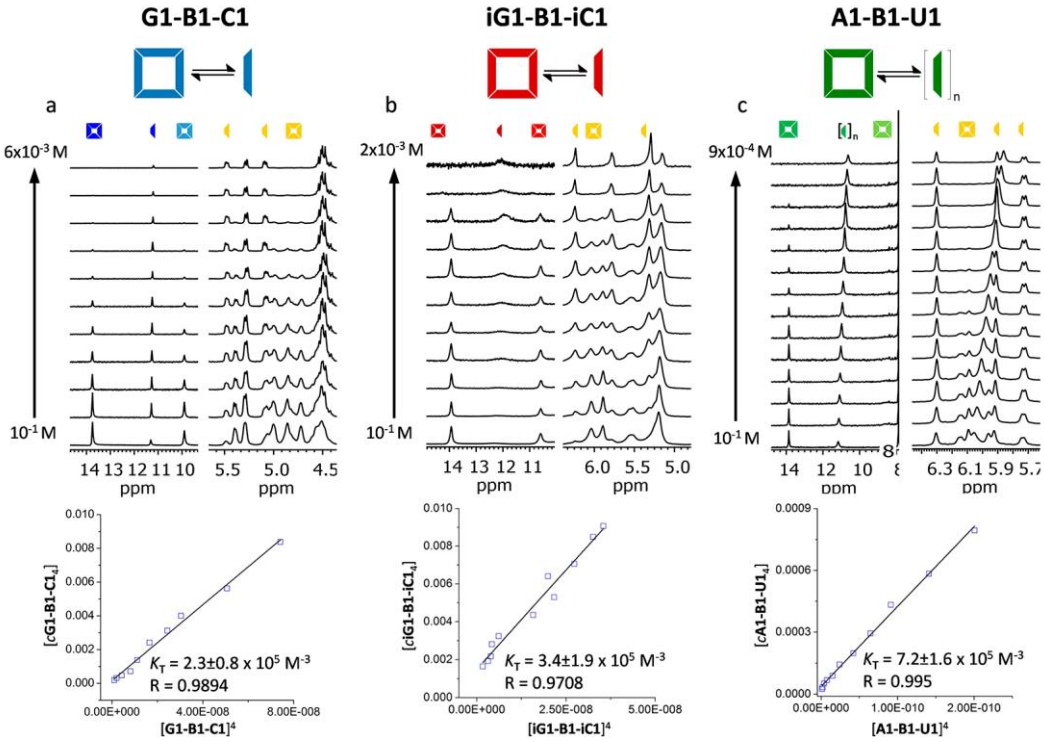


Figure 3.12. *Tetramerization Constants.* Changes in the H-bonded and ribose proton region (yellow) of the ^{1}H NMR spectra of (**a**) *c***G1-B1-C1**$_4$ and **(b)** *c***iG1-B1-iC1**$_4$ in pure DMF-d_7 and (**c**) *c***A1-B1-U1**$_4$ in 5:1 *v/v* $CDCl_3$-acetone-d_6 as a function of the concentration. Plots (**a**) [*c***G1-B1-C1**$_4$] vs [**G1-B1-C1**]^4 (**b**) [*c***iG1-B1-iC1**$_4$] vs [**iG1-B1-iC1**]^4 in pure DMF-d_7 and (**c**) [*c***A1-B1-U1**$_4$] vs [**A1-B1-U1**]^4 in 5:1 *v/v* $CDCl_3$-acetone-d_6. In all cases T = 298 K.

The analysis of these DMF-d_7 solutions by DOSY NMR (Figure 3.13) revealed now the presence of two species in slow exchange: the *c***G1-B1-C1**$_4$/*c***iG1-B1-iC1**$_4$ tetramer, having a lower diffusion coefficient, and the corresponding monomer.

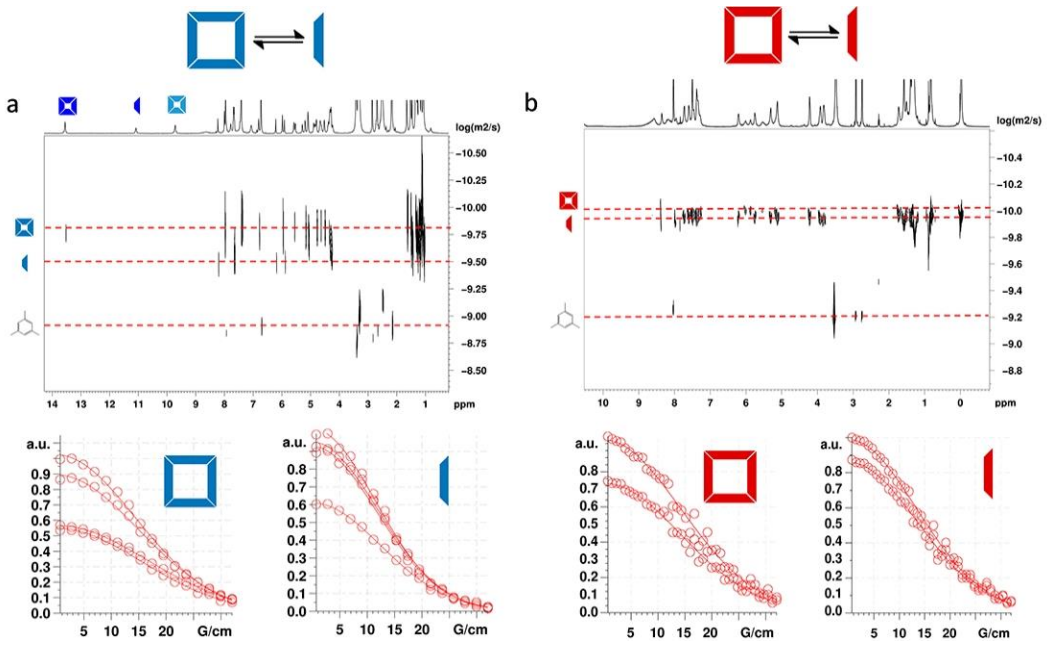


Figure 3.13. 2D-DOSY NMR spectrum of (**a**) **G1-B1-C1** and (**b**) **iG1-B1-iC1** with mesytilene as internal reference in DMF-d_7. (C = 1.0 x 10^{-2} M, T = 298 K).

Kinetic Analysis. EXSY ^{1}H NMR.

EXSY NMR experiment is a useful tool to assess the slow-exchange equilibria between monomer and tetramer. For such purpose, is necessary carry out a NOESY spectrum at τ_m = 0 ms and then, a T-ROESY in a suitable mixing time where the cross-peaks intensities between both species are almost equal Some ribose signals were considered appropriate to calculate the exchange rate constants, since they are well-separated and correspond to non-exchangeable C-H protons. Finally, EXSY NMR measurements (Figure 3.14) confirmed that the exchange between these two chemical species is remarkably slow for such a polar solvent, and exchange rate constants of k = 3.0 ± 0.7 s^{-1} (*c***G1-B1-C1**$_4$) and k = 7.2 ± 0.2 s^{-1} (*c***iG1-B1-iC1**$_4$) were calculated. The analysis of *c***A1-B1-U1**$_4$ by EXSY NMR in $CDCl_3$ afforded an exchange rate constant between the two slowly exchanging species of k = 6.6 ± 0.3 s^{-1} (Table 3.2).

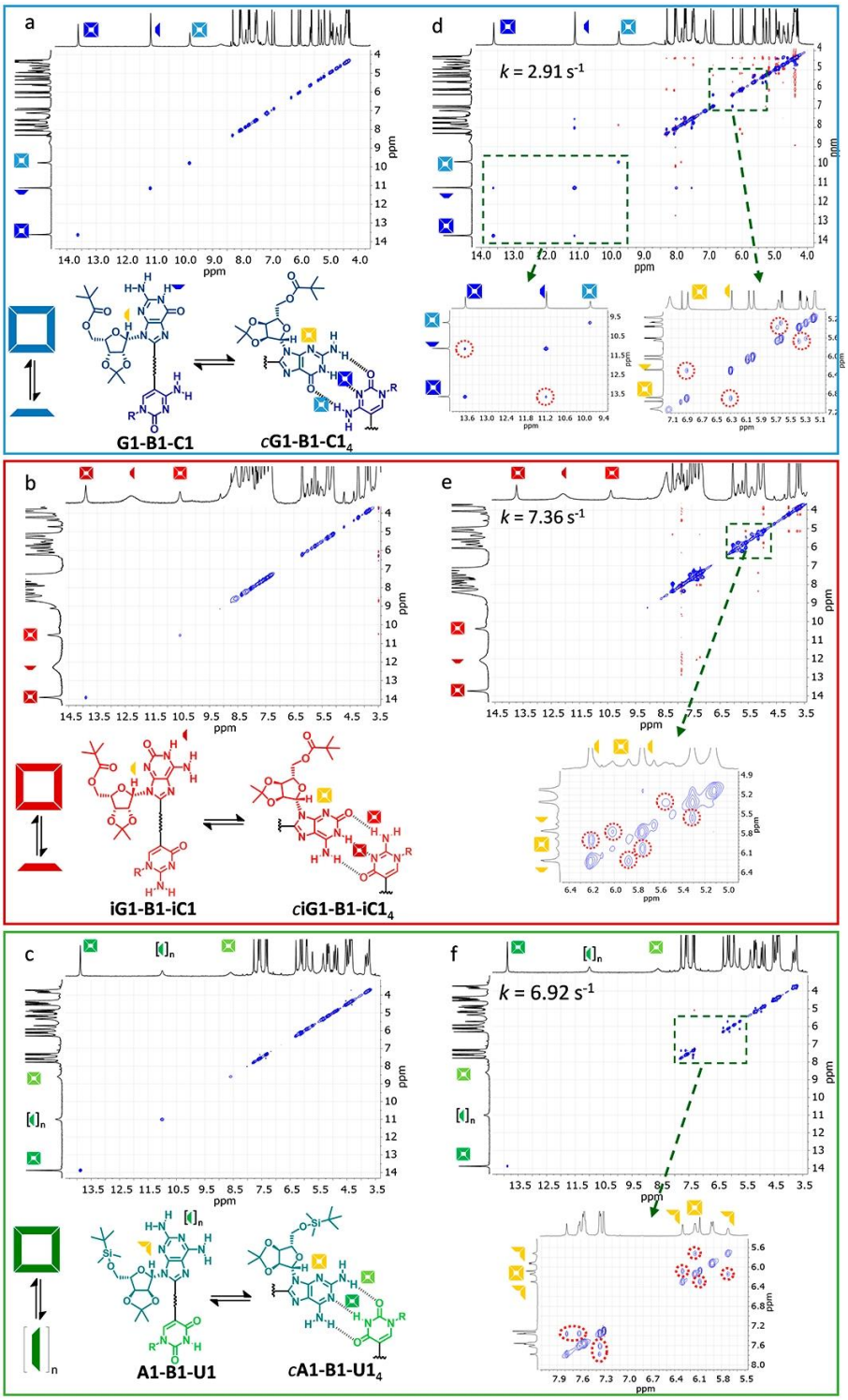


Figure 3.14. Selected region of the (**a**) NOESY spectrum of *c***G1-B1-C1**$_4$ (τ_m = 0 ms), (**b**) *ci***G1-B1-iC1**$_4$ (τ_m = 0 ms) in DMF-d_7 and (**c**) *c***A1-B1-U1**$_4$ (τ_m = 0 ms) in $CDCl_3$. T-ROESY spectrum of (**d**) *c***G1-B1-C1**$_4$ (τ_m = 200 ms), (**e**) *ci***G1-B1-iC1**$_4$ (τ_m = 150 ms) in DMF-d_7 and (**f**) *c***A1-B1-U1**$_4$ (τ_m = 150 ms) in $CDCl_3$. The ribose proton region is magnified for each spectrum. In all cases, C = 1.5x10^{-2} M, T = 298 K.

The exchange constant values have been calculated by two different methods:

a) Using the equations shown below, where *k* is the exchange rate constant, τ_m is the mixing time, X_A and X_B are the molar fractions of molecules in states A and B, respectively, I_{AA} and I_{BB} are the diagonal peak intensities, and I_{AB} and I_{BA} are the cross-peak intensities, we obtained values for *k*, which are the sum of the forward (association; k_1) and backward (dissociation; k_{-1}) pseudo-first order rate constants for the assembly process.

$$k = \frac{1}{\tau_m}\ln\frac{r+1}{r-1} \qquad r = 4X_AX_B\frac{I_{AA}+I_{BB}}{I_{AB}+I_{BA}} - (X_A - X_B)^2$$

b) Using the software EXSY Calc (from MestreLab Research, available at http://mestrelab.com/software/), which affords a quantitative analysis of the experimental intensities of the NMR peaks obtained in EXSY experiments to calculate the magnetization exchange rates of the exchange equilibrium. EXSY Calc directly calculates the forward (association; k_1) and backward (dissociation; k_{-1}) pseudo-first order rate constants by resolving the corresponding exchange rate matrix. Then, $k = k_1 + k_{-1}$.

M	τ_m ms	Method a *k* s^{-1}	Method b *k* s^{-1}
G1-B1-C1[a]	200	**2.46**	**2.76**
	100	**3.70**	**3.69**
	50	**2.35**	**2.30**
iG1-B1-iC1	150	**7.06**	**7.36**
A1-B1-C1	150	**6.33**	**6.92**

[a] Reference 1b. An average value of $k = 3.0 \pm 0.7$ s^{-1} was taken for the **G1-B1-C1** monomer-tetramer exchange rate constant in DMF-d_7.

Table 3.2. Kinetic data of *c***G1-B1-C1**$_4$, *c***iG1-B1-iC1**$_4$ and *c***A1-B1-U1**$_4$ ($C = 1.0\text{x}10^{-2}$, $T = 298$K) obtained by both methods.

Thermodynamic Analysis. Van't Hoff Plots.

Finally, increasing the temperature of a **G1-B1-C1** solution in highly polar solvents (DMF-d_7 and 1:1 *v/v* $CDCl_3$-DMSO-d_6)[161] resulted in tetramer dissociation to yield monomeric species (Figure 3.15). Once more, the shape and position of the G-amide and C-amine protons do not change significantly with temperature, indicating again a very slow exchange in the NMR timescale (even at high temperatures) and the sole presence of the **G1-B1-C1** and *c***G1-B1-C1**$_4$ species. The concentrations of **G1-B1-C1** and *c***G1-B1-C1**$_4$ were calculated in each spectrum by signal integration (at least 3 C-H proton signals for each species were averaged) and ln*K* was plotted *vs* T^{-1} (Van´t Hoff plot), yielding ΔH and ΔS values in each solvent system:

$$\ln(K) = -\frac{\Delta H^0}{R}\left[\frac{1}{T}\right] + \frac{\Delta S^0}{R}$$

Unfortunately, we could not calculate ΔH and ΔS values of **iG1-B1-iC1** and **A1-B1-U1** through this technique.

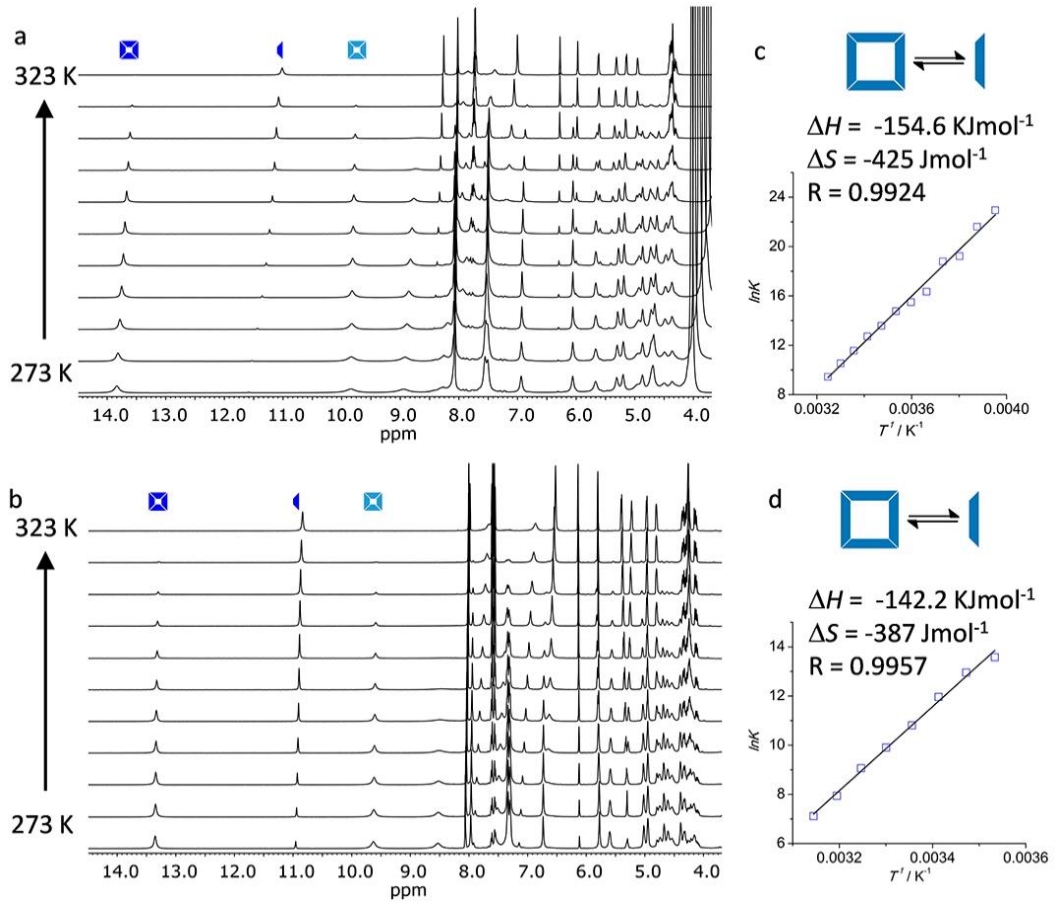


Figure 3.15. *Van't Hoff Plots.* 14.5-4.0 ppm region of the ^{1}H NMR spectra of **G1-B1-C1** in (**a**) pure DMF-d_7 and (**b**) 1:1 *v/v* $CDCl_3$-DMSO-d_6 solvent mixture as a function of the temperature (C = 1.0x10^{-2} M). (**c**, **d**) Van´t Hoff analysis of the temperature dependent data in each solvent system.

3.1.3 Study of the Self-Assembly Process by Optical Spectroscopy.

Below will be displayed a brief section with the most important results carried out by Jorge Camacho Garcia through optical spectroscopy measurements, which has been developed at the same time that this one. We though appropriate include these results here because of is a highly sensitive technique and allow us to assess the stability of our systems in organic solvents at high dilutions.

A second method we devised to study the cyclotetramerization equilibria is to employ low concentrations (10^{-3} to 10^{-7} M) and highly sensitive techniques, such as absorption, emission and circular dichroism (CD) spectroscopy. In these experiments, as it is described below, the choice of the solvent was very relevant again.

For **G1-B1-C1** and **iG1-B1-iC1**, DMF or DMAC (dimethylacetamide) solvents are too polar and, as previously observed in ^{1}H NMR dilution experiments (see Figure 3.12), the monomer is the only species present below a concentration of 10^{-3} M. At the other extreme, in apolar solvents that do not compete strongly for H-bonding, like CCl_4, $CHCl_3$, chlorobenzene (Figure 3.16 and Figure 3.17), or toluene, the tetramer is too stable to be dissociated

significantly by concentration or temperature changes.[162] Comparing monomer and tetramer spectroscopic features, the *c***G1-B1-C1**$_4$/*ci***G1-B1-iC1**$_4$ species are respectively characterized by an absorption shoulder at 393/401 nm, a red-shifted and low intensity emission maxima, and a Cotton effect with maxima at 357/366 and minima at 299/392 and 404/409 nm. We confirmed that such Cotton effect is only originated by cyclic H-bonding assembly, and not by $\pi-\pi$ stacking interactions or by single nucleoside interactions. The **G**+**C** or **iG**+**iC** 1:1 complexes, on the other hand, are CD-inactive when associated.[133] In solvents of intermediate polarity, like THF or dioxane, we could study the cyclotetramerization equilibria of **G1-B1-C1** and **iG1-B1-iC1** within the 2 x 10^{-4} – 1 x 10^{-6} concentration regime (Figure 3.16b and Figure 3.17b). At high temperatures or low concentrations the monomer species is dominant, while at low temperatures or high concentrations, the monomers are mostly associated as cyclic tetramers. The analysis of these data by concentration- and temperature-dependent experiments afforded the equilibrium constant (K_T) and the enthalpic (ΔH) and entropic (ΔS) changes associated with the cyclotetramerization process of **G1-B1-C1** and **iG1-B1-iC1** in THF, which are listed in Table 3.3.(see page 133). More information about how these calculations were made can be found in the corresponding published work.[158]

[162] a) It should be noted that the small spectroscopic changes observed by changing temperature (that is, a small red shift and an increase in intensity in absorption and emission when decreasing temperature) are common in oligo(phenyleneethynylene) and oligo(phenylenevinylene) molecules (see for instance DMAC where the monomer is the only one structure in solution and CCl_4 where the presence of the tetramer is quantitative) and are due to planarization of the π-conjugated system at low temperatures. These changes are obviously not observed in concentration-dependent measurements or in competition experiments at constant temperature; b) P. Jonkheijm, P. v.d. Schoot, A.P.H.J. Schenning E.W. Meijer, *Science,* **2006**, *313*, 80–83; c) D. González-Rodríguez, P. G. A. Janssen, R. Martín-Rapún, I. De Cat, S. De Feyter, A. P. H. J. Schenning, E. W. Meijer, *J. Am. Chem. Soc.* **2010**, *132*, 4710–4719.

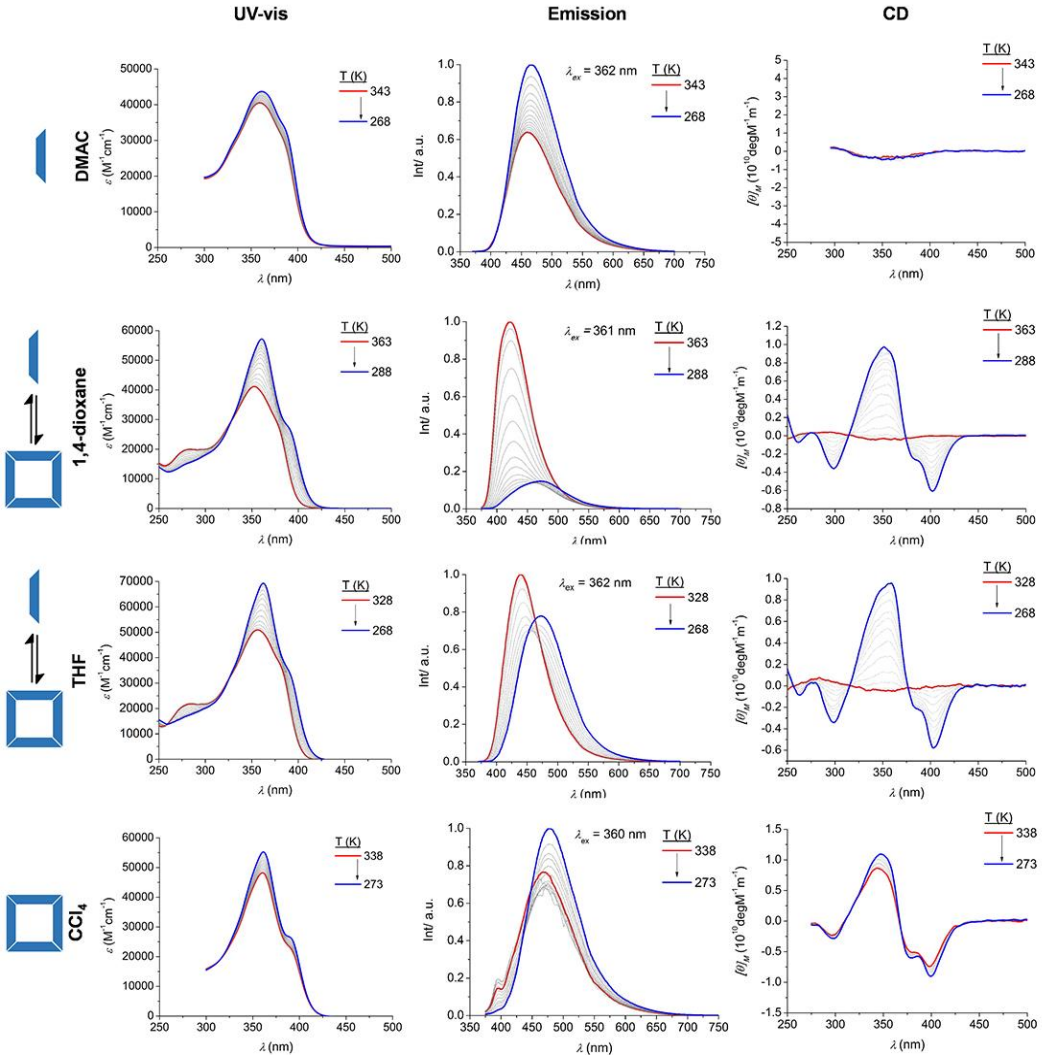


Figure 3.16. Temperature-dependent UV-vis (left), emission (center) and CD (right) spectroscopic changes of **G1-B1-C1** in different solvents (dimethylacetamide (DMAC), C = 1.0x10^{-4} M; 1,4-dioxane, C = 5.0x10^{-5} M; THF, C = 1.25x10^{-5} M; CCl_4, C = 1.25x10^{-5} M).

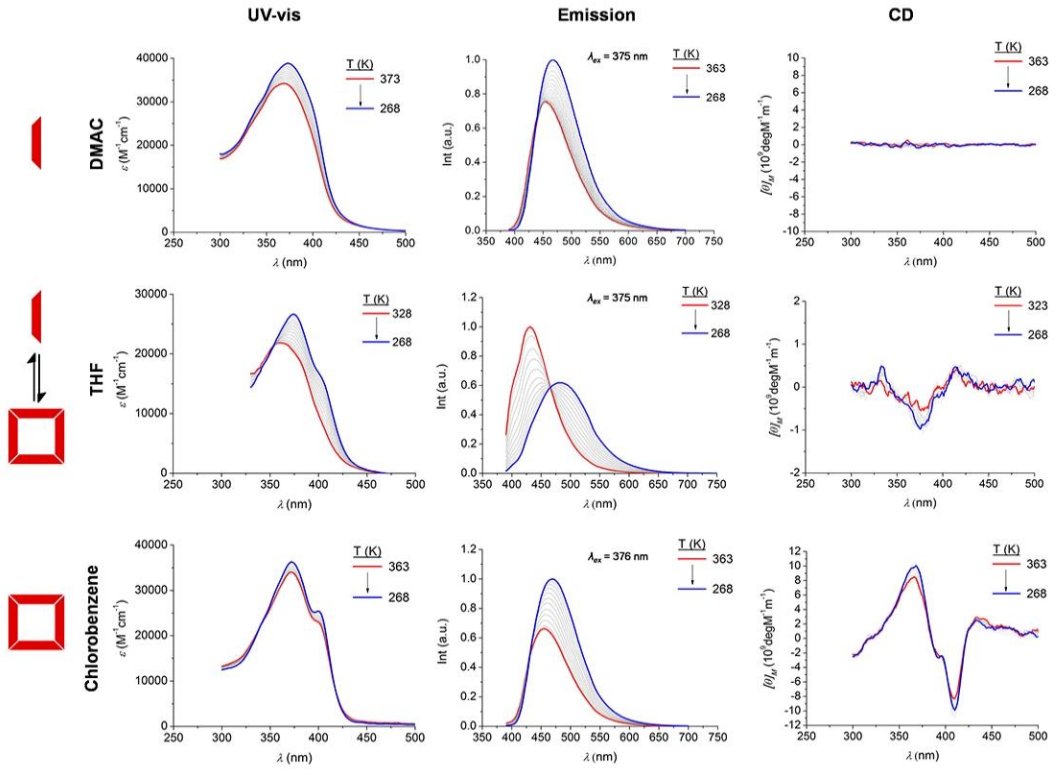


Figure 3.17. Temperature-dependent UV-vis (left), emission (center) and CD (right) spectroscopic changes of **iG1-B1-iC1** in different solvents (dimethylacetamide (DMAC), C = 2.0x10^{-4} M; THF, C = 1.0x10^{-5} M; Chlorobenzene, C = 3.0x10^{-4} M).

The situation is again different for monomer **A1-B1-U1**. Solvents like THF or dioxane are already too polar for this dinucleoside, which is present in solution only as monomeric species in the whole concentration and temperature range studied. The monomer-tetramer equilibria could only be investigated at these low concentrations in solvents of very low polarity, such as the 2:3 *v/v* $CHCl_3$-CCl_4 mixture, chlorobenzene or toluene (Figure 3.18). The *c***A1-B1-U1**$_4$ assembly displayed spectroscopic characteristics that are related to the other cyclic tetramers: red-shifted absorption and emission maxima, and a Cotton effect with maxima at 413 and minima at 371. Similar concentration- and temperature-dependent experiments as those carried out with **G1-B1-C1** and **iG1-B1-iC1** in THF were performed with **A1-B1-U1** in these solvents. The spectroscopic trends in the 2:3 *v/v* $CHCl_3$-CCl_4 mixture were fitted to appropriate models in order to calculate K_T, ΔH and ΔS (see Table 3.2. in page 124).[163]

[163] C. A. Hunter, M. C. Misuraca, S. M. Turega, *J. Am. Chem. Soc.* **2011**, *133*, 20416–20425.

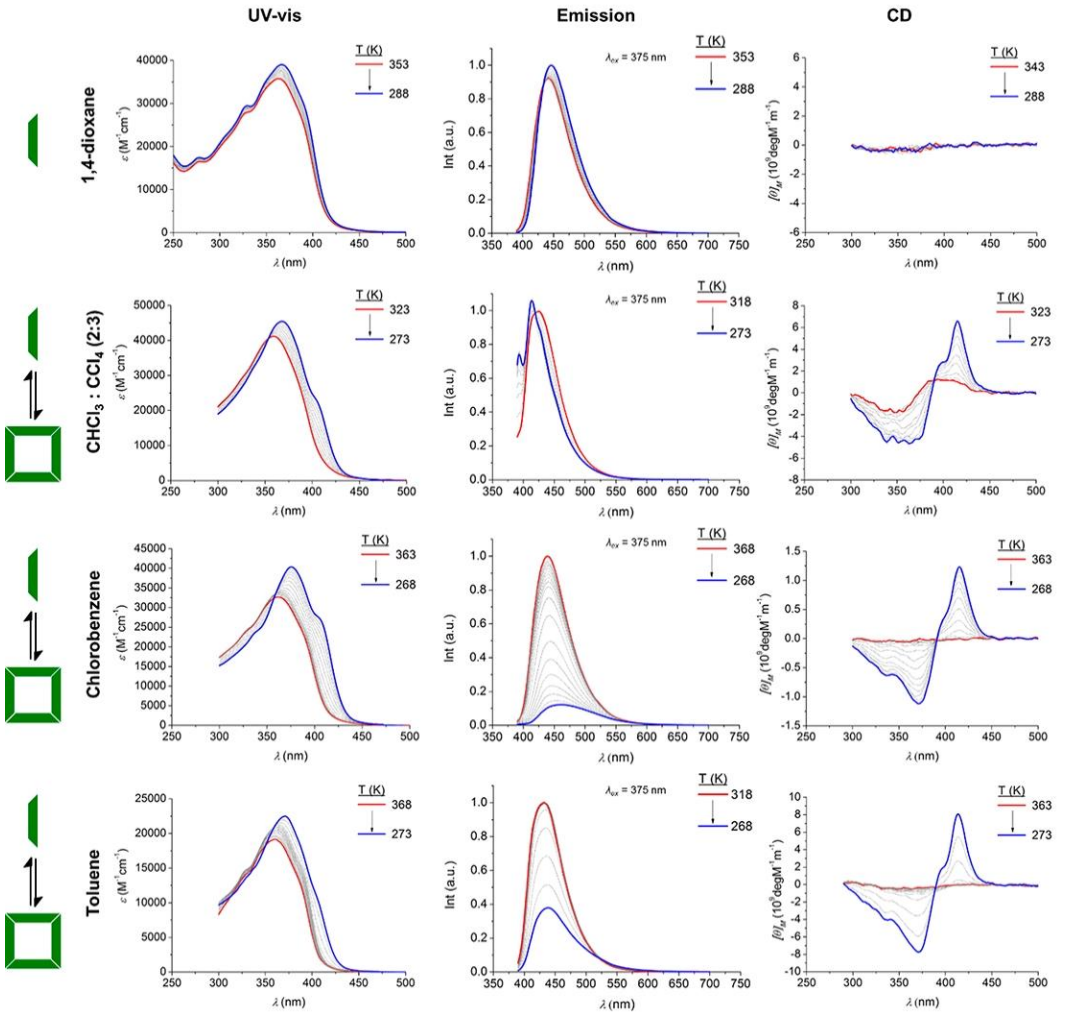


Figure 3.18. Temperature-dependent UV-vis (left), emission (center) and CD (right) spectroscopic changes of **A1-B1-U1** in different solvents (1,4-dioxane, C = 3.0x10^{-4} M; 2:3 *v/v* $CHCl_3$-CCl_4, C = 3.0x10^{-4} M; Chlorobenzene, C = 3.0x10^{-4} M, Toluene, C = 2.0x10^{-5} M).

3.1.4.Self-assembly into Cyclic Tetramers Studied by Denaturation Experiments.

Due to the impressive thermodynamic and kinetic stability exhibited by our cyclic tetramers, particularly by those formed from the **G1-B1-C1** and **iG1-B1-iC1** complementary dinucleosides, competition experiments were designed to indirectly evaluate the association constants of the cyclotetramerization process in an additional manner. In these experiments, increasing amounts of a complementary mononucleoside are gradually added to a solution of the associated tetramers in solvents of low polarity such as $CDCl_3$ or 2:3 *v/v* $CDCl_3$-CCl_4. The titration process was monitored by two different techniques: ^{1}H NMR at high concentrations (*ca.* 10^{-2} M) and emission spectroscopy at relatively low concentrations (below 10^{-4} M).

The addition of the corresponding pyrimidine mononucleoside **C/iC/U** (scheme at the top of Figure 3.19) to a solution of **G1-B1-C1/iG1-B1-iC1/A1-B1-U1** resulted in the gradual disappearance of the characteristic cyclic tetramer ^{1}H NMR signals and the emergence of a new set of signals attributed to the **C·G1-B1-C1/iC·iG1-B1-iC1/U·A1-B1-U1** complexes (Figure

3.19). Due to the high concentrations reached and the broadening of the ^{1}H NMR signals during the titration of *c***iG1-B1-iC1**$_4$ in $CDCl_3$, we could only estimate how many equivalents of **iC** have to be added in order to dissociate the cyclic tetramer completely, but we were no able to calculate the association constant. For this reason and the similar thermodynamic and kinetic stability with **G1-B1-C1** we decided perform the following denaturation experiments with this last one monomer. Hence, optical spectroscopic techniques were used to assess the behaviour of *c***iG1-B1-iC1**$_4$ through this competition method.

Denaturation experiments revealed different stabilities between monomers **G1-B1-C1** and **A1-B1-U1**. On the one hand, the complex **C·G1-B1-C1** are in fast exchange with excess **C** and this species is, at the same time, in slow exchange with *c***G1-B1-C1**$_4$ in $CDCl_3$ (where the G-H amide is well resolved) and THF-d_8. When DMF-d_7, a more polar solvent is employed, the species corresponding to **C·G1-B1-C1** and the excess of **C** are now in fast equilibrium with the **G1-B1-C1** monomer since the binding constant K_{ref} is much smaller in this solvent. This is clearly observed in the G-H amide signal shifting from 11 ppm to 13 ppm as the concentration of **C** increases (see Figure 19c). Is worthy to mention that the same trend is observed in the titration experiments on the reference compounds **G** with **C** commented previously in *Chapter 2*. On the other hand, the denaturation experiments of the weaker *c***A1-B1-U1**$_4$ (as occurred with *c***G1-B1-C1**$_4$ in $CDCl_3$ and THF) showed that the complex **U·A1-B1-U1** are in fast exchange with excess **U** and this species is, at the same time, in slow exchange with *c***A1-B1-U1**$_4$ in $CDCl_3$ (in which the tetramer is not formed quantitatively) and in the less polar 2:3 *v/v* $CDCl_3$-CCl_4 solvent mixture leading both solvents to a quantitatively similar outcome (Figure 19d,e).

From the analysis of these equilibrium concentrations at different **C**/**iG**/**U** concentrations the equilibrium constant (K_C) of the competition experiment could be obtained and, from here, the K_T and the *EM* values, as will be explained later. However, this was not always possible due to practical reasons. For instance, in ^{1}H NMR it was not always possible to find several isolated C-H signals that offered a reliable integration, as happened with**·iG1-B1-iC1** and **A1-B1-U1** in $CDCl_3$. In addition, the data obtained from the relative concentrations of *c***A1-B1-U1**$_4$ and **U·A1-B1-U1** in 2:3 *v/v* $CDCl_3$-CCl_4 could not be fitted appropriately by our model, probably because more complex equilibria are taking place in this system involving small open oligomers.

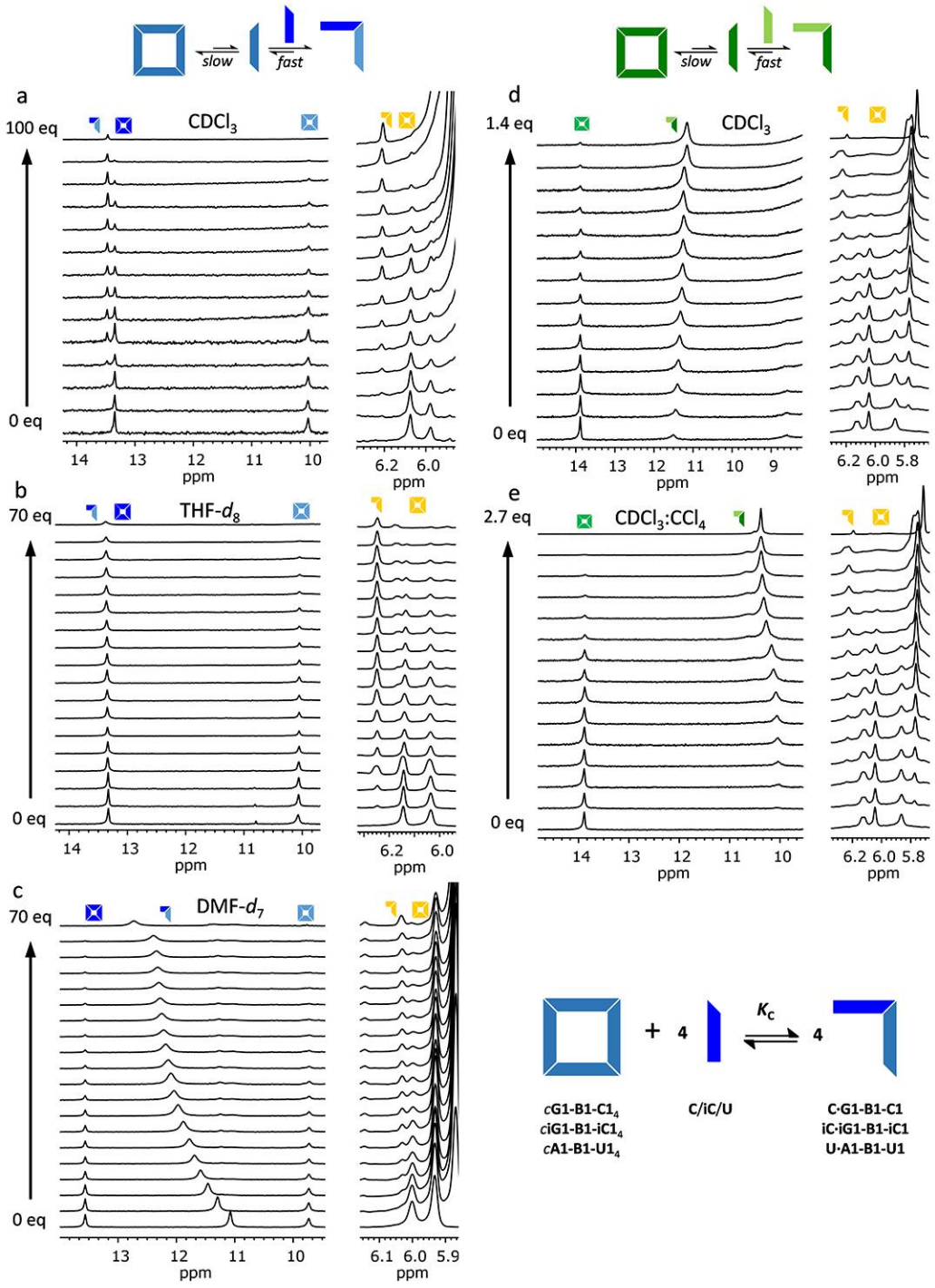


Figure 3.19. ^{1}H NMR titration experiments of (**a**) **G1-B1-C1** with **C** in $CDCl_3$, THF-d_8 and DMF-d_6 (**b**) **iG1-B1-iC1** with **C** in $CDCl_3$ (**c**) **A1-B1-U1** with **U** in 2:3 *v/v* $CDCl_3$-CCl_4. In all cases: C = 1x10^{-3} M, T = 298 K.

Absorption and CD spectroscopy could not be employed in these competition measurements due to signal saturation after surpassing a few equivalents of the titrating mononucleoside molecule. Contrarily, emission spectroscopy proved useful here, since **G1-B1-**

C1/**iG1-B1-iC1**/**A1-B1-U1** could be excited selectively within the 365-400 nm range, a region where the **C**/**iC**/**U** nucleosides do not absorb. Hence, the recorded fluorescence spectra correspond only to the **G1-B1-C1**/**iG1-B1-iC1**/**A1-B1-U1** molecules in different association states. As *c***G1-B1-C1**$_4$/*c***iG1-B1-iC1**$_4$/*c***A1-B1-U1**$_4$ is dissociated with excess **C**/**iC**/**U**, respectively, a blue emission shift was noted in all cases (Figure 3.20), which resembles the changes previously monitored by decreasing concentration or increasing temperature. The competition experiments with *c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$ were performed in $CHCl_3$, but dissociation of *c***A1-B1-U1**$_4$ was monitored again in the 2:3 *v/v* $CDCl_3$-CCl_4 mixture, since $CHCl_3$ solutions at low concentrations contained basically monomeric species (see Figure 3.20).

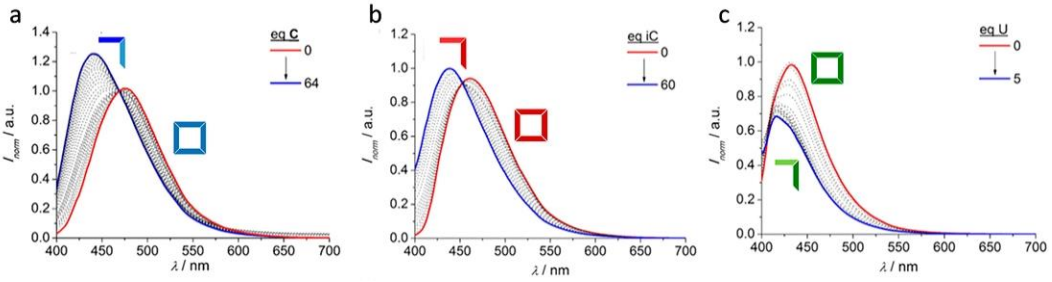


Figure 3.20. Fluorescence emission changes observed in the titration of (**a**) **G1-B1-C1** with **C** in $CHCl_3$ (*C* = 5x10^{-5} M, 298 K, λ_{exc} = 390 nm) (**b**) **iG1-B1-iC1** with **iC** in $CHCl_3$ (*C* = 1x10^{-5} M, 298 K, λ_{exc} = 375 nm) and (**c**) **A1-B1-U1** with **U** in 2:3 *v/v* $CDCl_3$-CCl_4 (*C* = 2x10^{-4} M, 298 K, λ_{exc} = 381 nm).

The degree of cyclic tetramer (*c***M**$_4$) association, defined as the fraction of dinucleoside (**M**) assembled as a cyclic tetramer at each titration point, could be calculated by direct ^{1}H signal integration, providing several suitable C-H signals for each species are available, or by analysis of the emission maximum trends. This magnitude is plotted in Figure3.21a as a function of the mononucleoside stopper equivalents added. Again, a clear difference towards dissociation was noted for *c***G1-B1-C1**$_4$ or *c***iG1-B1-iC1**$_4$, on one hand, and for *c***A1-B1-U1**$_4$, on the other. While *c***G1-B1-C1**$_4$/*c***iG1-B1-iC1**$_4$ could resist up to 60 equivalents of **C**/**iC** mononucleosides (even in THF-d_8 or DMF-d_7), *c***A1-B1-U1**$_4$ was fully dissociated after the addition of *ca.* 3 equivalents of **U**, independently of the solvent system employed.

Actually, in these experiments the intramolecular and intermolecular base pair binding events are made to compete. As a matter of fact, as depicted in the scheme of Figure 3.21b, the *EM* of the system can be inversely related to the competition equilibrium constant (K_C; equilibrium 1) as: $EM = 1/K_C$, making the approximation that K_2 (equilibrium 2) equals the reference base pair binding constant (K_{ref}; equilibrium 3), which was obtained in each solvent in separate titration experiments as is described in *Chapter 2*. In short, these relevant competition experiments directly suggest that the lower stability of *c***A1-B1-U1**$_4$ *not only comes from a substantial decrease in the individual nucleobase association constants* (K_{ref}), *but also chelate cooperativity is much weaker in this system.*

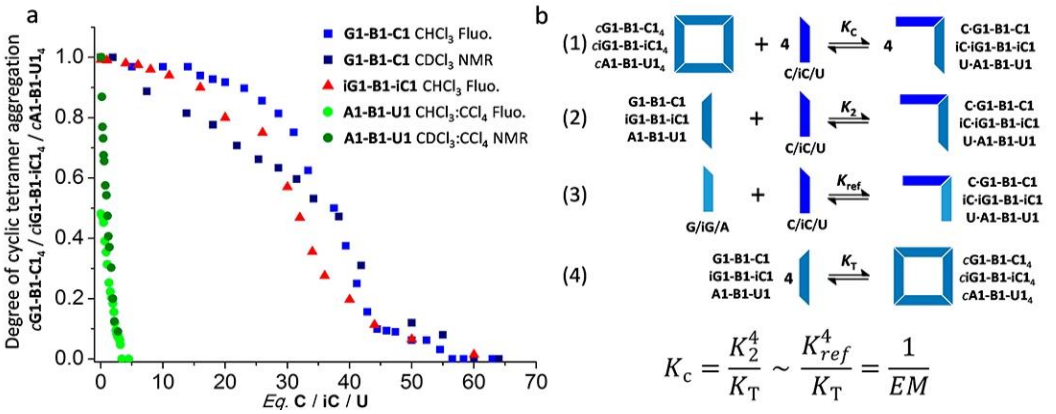


Figure 3.21 (**a**) Plots of the degree *c***G1-B1-C1**$_4$, *c***iG1-B1-iC1**$_4$ and *c***A1-B1-U1**$_4$ association, measured by either ^{1}H NMR or emission spectroscopy, as a function of the equivalents of **C**/**iC**/**U** added. (**b**) Representation of the different equilibria taking place in the competition experiments.

Estimation of *EM* values for *c*G1-B1-C1$_4$, *c*iG1-B1-iC1$_4$ and *c*A1-B1-U1$_4$ and Speciation Curves.

The whole set of results exposed so far and the thermodynamic data obtained from diverse experiments in different conditions clearly demonstrate that **A1-B1-U1** forms considerably less stable cyclic tetramers than **G1-B1-C1** or **iG1-B1-iC1**, which, on the other hand, reveal comparable qualitative and quantitative association behaviour (Table 3.3).

M	Solvent	K_T M^{-3}	K_{ref}[a] M^{-1}	*EM* M
*c***G1-B1-C1**$_4$	DMF	2.3±0.8x10^5 [b]	5.7±0.3	218
	THF	9.1±4.0x10^{14} [c]	1.5±0.1x10^3	180
		3.7±0.3x10^{15} [d]		730
	$CHCl_3$	5.6±3.1x10^{20} [e]	2.8±0.3x10^4	910
		5.0±0.1x10^{20} [f]		813
*c***iG1-B1-iC1**$_4$	DMF	3.4±1.9x10^5 [b]	6.1±0.8	246
	THF	3.7±1.2x10^{15} [c]	1.7±0.6x10^3	463
		2.2±0.5x10^{15} [d]		294
	$CHCl_3$	3.3±0.4x10^{20} [f]	3.2±0.5x10^4	314
*c***A1-B1-U1**$_4$	$CHCl_3$		2.5±0.4x10^2	0.10[h]
	$CHCl_3$-CCl_4 (2:3)	9.4±0.3x10^{11} [c]	2.0±0.4x10^3	0.06
		2.8±0.2 x10^{11} [d]		0.02
	$CHCl_3$-acetone(5:1)	7.2±1.6 x10^6 [g]	0,9±0.6 x10^2	0.11

Table 3.3. Cyclotetramerization constants (K_T), reference intermolecular association constants (K_{ref}) and effective molarities (*EM*) obtained for **G1-B1-C1**/**iG1-B1-iC1**/**A1-B1-U1** from different experiments. [a]Determined from titration experiments with the mononucleosides: **G**+**C**, **iG**+**iC**, **A**+**U**.[133] Data calculated from: [b] ^{1}H NMR dilution. [c] UV-vis dilution. [d] Temperature-dependent experiments. [e] ^{1}H NMR competition. [f] Fluorescence competition. [g] ^{1}H NMR dilution. [h] Estimated from the fitting of the ^{1}H NMR dilution data.

This trend was of course expected, since the individual **A**:**U** binding constant (K_{ref} ~ 2.5x10^2 M^{-1}) is typically about 2 orders of magnitude lower than **G**:**C** or **iG**:**iC** (K_{ref} ~ 3x10^4 M^{-1}) in $CHCl_3$ (see *Chapter 2* and Table 3.3).[133] This can be explained by the Jorgensen model[43] and is due to the different stabilizing/destabilizing secondary H-bonding interactions between vicinal donor and acceptor groups in the *DAD-ADA* (**A**:**U**) pair *versus* the *DDA-ADD* (**G**:**C** or **iG**:**iC**) H-bonding pair. Since $K_T = K_{ref}^4\ EM$, this difference should lead to a reduction in the **A1-B1-U1**

cyclotetramerization constants of about 10^8 M^{-3}, when compared to **G1-B1-C1** or **iG1-B1-iC1**, considering that *EM* is independent of the binding interaction and similar for the three systems. However, the experimental results obtained in this work and exposed in Table3.3, suggest that such reduction is actually much larger. In fact, our results indicate that K_T for **A1-B1-U1** is not only reduced by a decrease in K_{ref}, but also by a substantial decrease in the magnitude of *EM*. The competition experiments shown in Figure 3.21, as explained above, clearly support this hypothesis.

The *EM* values for each dinucleoside cyclotetramerization process may be estimated from the K_T values obtained in the different experiments presented and the corresponding reference association constants (K_{ref}), previously determined by titration experiments with the mononucleosides (*i.e.* **G**+**C**, **iG**+**iC**, **A**+**U**) in *Chapter 2*,[133] using the relationship: $EM = K_T / K_{ref}^4$. These K_{ref} and *EM* values were used to simulate speciation curves for each dinucleoside molecule in different solvent systems. These curves, which relate the concentration of each supramolecular species with total concentration, are able to reproduce quite satisfactorily the dissociation behavior observed for our *c***G1-B1-C1**$_4$/*ci***G1-B1-iC1**$_4$/*c***A1-B1-U1**$_4$ tetramers in dilution experiments in different solvents. Both simulated and experimental results are combined in Figure 3.22a, for *c***G1-B1-C1**$_4$/*ci***G1-B1-iC1**$_4$ in $CHCl_3$, THF, and DMF, and Figure 3.22b, for *c***A1-B1-U1**$_4$ in 2:3 *v/v* $CHCl_3$-CCl_4, $CHCl_3$, and 5:1 *v/v* $CHCl_3$-acetone. Is interesting to note that the obtaining of these values have a high error due to the difficulty to integrate properly each ^{1}H NMR signal along the each titration. Hence, these speciation curves are very useful for us in order to estimate the final EM values.

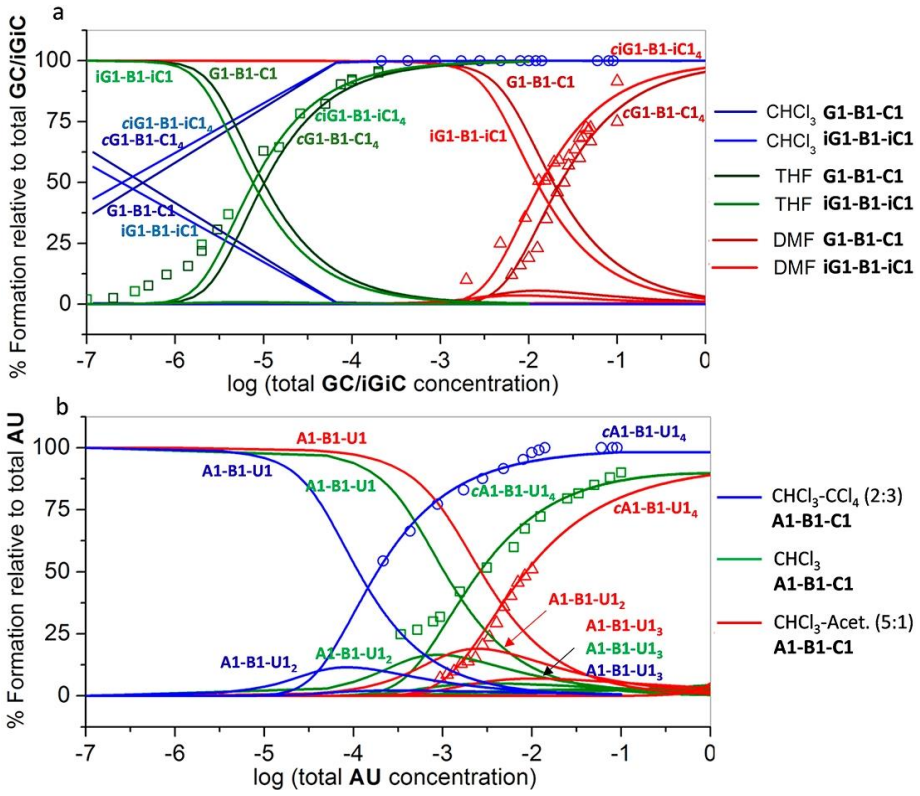


Figure 3.22. Simulated speciation curves (lines) and experimental dilution data indicating the degree of c**M**$_4$ association of (**a**) **G1-B1-C1**/**iG1-B1-iC1** in $CHCl_3$ (circles), THF (squares) and DMF (triangles), and (**b**) **A1-B1-U1** in 2:3 *v/v* $CHCl_3$-CCl_4 (circles), $CHCl_3$ (squares), and 5:1 *v/v* $CHCl_3$-acetone (triangles).

The dissociation of *c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$ was reproduced using the experimentally determined K_{ref} values and *EM* values that range between 200 and 700 M, which is in good agreement with the calculated *EM*s in this Chapter. In all cases, *c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$ fulfill the condition K_{ref}·*EM* > 185·*n*, (*n* being the number of monomers in the cycle; *n* = 4), defined by Ercolani to reach complete cycle assembly at a certain monomer concentration. However, the concentration range where this condition is met is clearly wider in solvents that do not compete strongly for H-bonding and thus maintain a high K_{ref} value. As a matter of fact, the lower self-assembly concentration (*lsac*), meaning the concentration at which half of the monomer is assembled into cycles, was approximately evaluated as: *lsac*(DMF) ~ 5 x 10^{-2} M, *lsac*(THF) ~ 10^{-5} M, and *lsac*($CHCl_3$) ~ 5 x 10^{-7} M.

lsac values are calculated through the principal factors of the macrocyclization process and can be obtained as follows:

$$\boldsymbol{lsac} = \frac{\mathbf{2}}{\boldsymbol{n}^{\mathbf{1}/(\boldsymbol{n}-\mathbf{1})}\, \boldsymbol{EM}_{\boldsymbol{n}}^{\mathbf{1}/(\boldsymbol{n}-\mathbf{1})}\, \boldsymbol{K}_{\boldsymbol{inter}}^{\boldsymbol{n}/(\boldsymbol{n}-\mathbf{1})}}$$

Where n correspond to the number of the monomers that constitute the macrocycle, *EM* is the ratio K_T/K_{ref} (K_{intra}/K_{inter}) and finally K_{inter} (K_{ref}) is the value of the association constant between two mononucleosides.

On the contrary, *EM* values between 0.05 and 0.1 had to be set to reproduce appropriately the dilution trends of *c***A1-B1-U1**$_4$ in the three solvent systems. This represents a reduction of more than 3 orders of magnitude in *EM* with respect to the **G1-B1-C1**/**iG1-B1-iC1** cyclotetramerization process. As a result, and in line with our experimental observations, the condition K_{ref}·*EM* > 185·*n* is hardly fulfilled by **A1-B1-U1** even in 2:3 *v/v* $CHCl_3$-CCl_4, meaning that quantitative cycle assembly could only be achieved in strongly apolar media where K_{ref} is enhanced. Furthermore, *lsac* values for *c***A1-B1-U1**$_4$ are considerably decreased with respect to *c***G1-B1-C1**$_4$/*c***iG1-B1-iC1**$_4$ in the same conditions: *lsac*($CHCl_3$) ~ 5 x 10^{-3} M.

It is interesting to note that related cyclic tetramers based on pyridine-metal coordination, a strong, single-point interaction that allows for some degree of torsional and rotational flexibility, afford *EM* values that are intermediate between our symmetric and unsymmetric patterns (*EM* = 0.1–20 M). We would also like to remark, as previous studies have also shown, that these supramolecular *EM*s do not seem to follow any clear relationship with solvent polarity.

Symmetric *versus* Unsymmetric Hydrogen-bonding Patterns.

Being a thermodynamic parameter, the *EM* has both an enthalpic and an entropic component. The enthalpic contribution of *EM* is mainly dominated by the creation of ring strain upon cyclization. Therefore, preorganization of the structure of the monomeric components towards a certain cycle size arises as a fundamental design factor to achieve high *EM*s and hence quantitative yields of the target self-assembled structure. On the other hand, entropy is responsible for the preference of an intramolecular binding event with respect to the intermolecular one, since in this way some rotational and translational molecular degrees of freedom are not lost upon association. However, cyclization also involves a loss of entropy due to restriction of certain degrees of freedom, such as torsional or rotational motions, in the closed species. This is more accused when the monomer has a flexible structure, with several rotatable bonds, and therefore rigid monomers typically increase the magnitude of *EM*.

In our case, all **G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-U1** monomers share a rigid structure that was designed to produce cyclic squared-shaped assemblies devoid of strain and with minimal conformational entropy loss. This was achieved, on one hand, profiting from the 90° angle that the 8-purine and 5-pyrimidine positions adopt upon Watson-Crick complementary base pairing, and, on the other, employing a rigid central block connecting the bases with only 4 rotatable linear π-conjugated bonds. As shown in Figure 23, rotation around these bonds can produce different conformations in which the Watson-Crick edges alternate between *syn* and *anti* relative arrangements in the monomer and open oligomers. However, cycle formation demands the Watson-Crick edges to be in a *syn* relative conformation (although rotation around these σ-bonds remains allowed in the cycle). This is a degree of freedom that is lost when comparing cyclic and open **G1-B1-C1**, **iG1-B1-iC1**, and **A1-B1-U1** oligomers, and must contribute to a reduction, of entropic origin, in the maximum attainable *EM* of the cyclic system.

Now, the **A1-B1-U1** monomer, having complementary nucleosides that pair with a symmetric DAD-ADA H-bonding pattern, have the possibility to self-assemble *via* either Watson-Crick or reverse Watson-Crick interactions (Figure 3.23). Each binding mode provides a different association angle (90° and 210°, respectively) and their relative energy is assumed to be comparable, as previous studies with the adenine-thymine pair have demonstrated. Furthermore, additional **A**:**U** binding modes could be considered that do not differ much in association strength from the Watson-Crick mode, such as the double H-bonding interaction of the U base with the 2-aminoadenine Hoogsten edge (see introduction). This introduces additional degrees of freedom, not available in the unsymmetric (non-rotatable) *ADD-DAA* or *DDA-AAD* patterns, which allow the linear **A1-B1-U1**$_n$ oligomers to access a higher number of binding and conformational possibilities. However, such freedom must be lost upon cycle formation because the cyclotetramerization process exclusively demands a 90° Watson-Crick interaction. Hence, the entropy loss associated to cyclization becomes much larger for *c***A1-B1-U1**$_4$ than for *c***G1-B1-C1**$_4$ or *c***iG1-B1-iC1**$_4$, resulting in a supplementary and notable reduction of the *EM* values.

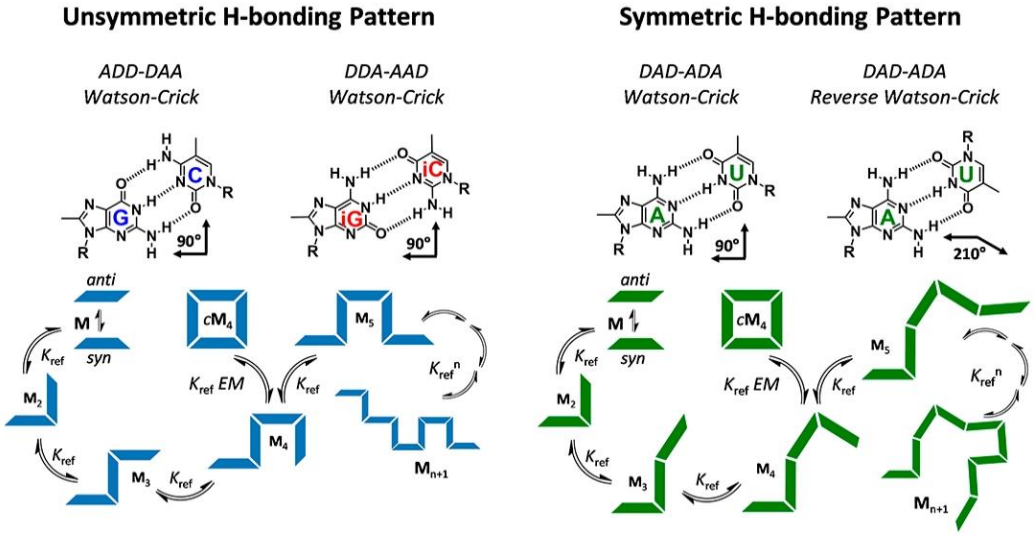


Figure 3.23. *Unsymmetric versus Symmetric H-bonding patterns.* Due to the ability of the AU monomer to bind either via Watson-Crick or Reverse Watson-Crick interactions, the H-bonded oligomeric species can access a higher number of conformational possibilities that must be lost upon cyclization.

It would have been highly interesting and desirable to compare the cyclization process of our **A1-B1-U1**, **G1-B1-C1** and **iG1-B1-iC1** molecules with a related dinucleoside having complementary *DDD-AAA* base pairs. However, a purine-pyrimidine couple having such H-bonding pattern and that would maintain the same geometric requirements of our cyclic tetramer system is, to the best of our knowledge, not chemically available. We however propose, given the results obtained in this work, that such cyclic tetramer bound by symmetric *DDD-AAA* pairs would be endowed with both a high K_{ref} (*ca.* 10^5–10^6 M in $CHCl_3$, given the data reported in the literature) and a low *EM* (0.01–0.1, comparable also to our H-bond symmetric ***c*A1-B1-U1**$_4$ system). From these values, we simulated in Figure 3.24 the speciation curves of the hypothetical symmetric *DDD–AAA* system in $CHCl_3$ and compared it with the ones obtained for unsymmetric *ADD-DAA*/*DDA-AAD* and symmetric *DAD-ADA* H-bonded systems. For the sake of clarity, only a few supramolecular species are represented in Figure 3.24 within the 10^{-8} to 10^8 M concentration range: **M**, **M**$_2$, **M**$_3$, **M**$_4$, ***c*M**$_4$, **M**$_5$, and, as representative examples of higher-order H-bonded linear oligomers: **M**$_{10}$, **M**$_{15}$, and **M**$_{20}$.

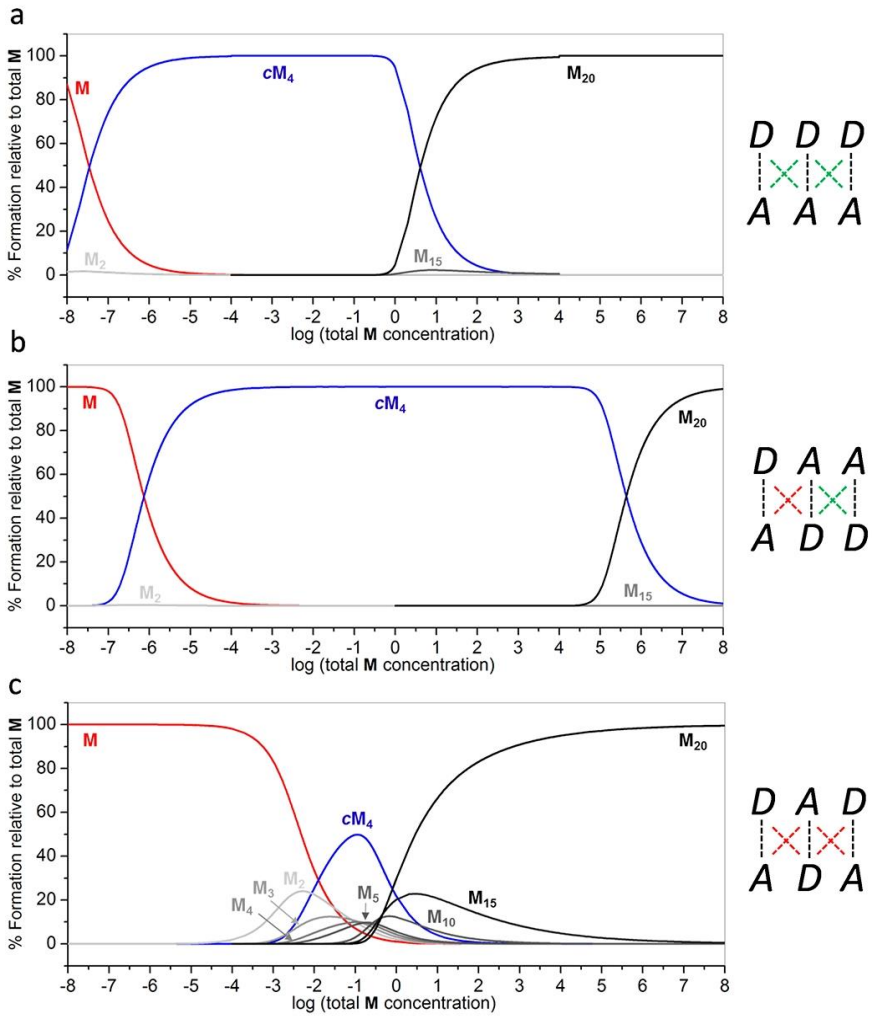


Figure 3.24 Simulated speciation curves for the H-bonded self-assembly of hypothetic (**a**) *DDD-AAA* (K_{ref} = 10^6 M^{-1}; EM = 0.05 M), (**b**) *ADD-DAA*/*DDA-AAD* (K_{ref} = 10^4 M^{-1}; *EM* = 500 M), and (**c**) *DAD-ADA* (K_{ref} = 10^2 M^{-1}; EM = 0.05 M) ditopic monomers with identical geometrical features to the ones studied in this work.

It is clear that the use of an unsymmetric *ADD-DAA*/*DDA-AAD* H-bonding pattern (Figure 3.24b), providing moderate K_{ref} and high *EM*s, leads to cyclic assemblies that persist as the main species in solution over a much broader concentration range. Only at very low concentrations $c\mathbf{M}_4$ is dissociated as a monomer, but no other associated species is seen to compete within the 10^{-8} to 10^{5} M concentration range, underlining the strong all-or-nothing behavior of this H-bonded system. At very high concentrations intra- and intermolecular processes begin to compete, and above 10^{5} M linear polymers of high molecular weight begin to grow at the expense of the $c\mathbf{M}_4$ species. On the other hand, the *DDD-AAA* pattern (Figure 3.24a) leads to more strongly bound assemblies along the whole concentration scale, as a result of a high K_{ref} value, and the monomer is only present at concentrations below 10^{-5} M. The $c\mathbf{M}_4$ species is dominant between the 10^{-5} - 1 M range but, in sharp contrast to the unsymmetric pattern, higher-order linear oligomers begin to compete strongly at moderate concentrations. Finally, in the weaker *DAD-ADA* H-bonded system (Figure 3.24c) no associated species is seen below 10^{-4} M. At intermediate concentrations a distribution of small oligomers (**M**, $\mathbf{M}_2$, $\mathbf{M}_3$, $\mathbf{M}_4$, $\mathbf{M}_5$,..), among which the cyclic $c\mathbf{M}_4$ is one of the main species, is observed. In analogy to the *DDD-AAA* system, increasing concentration above 1 M makes the higher-order linear oligomers to dominate. However, high molecular weight distributions are much faster at lower concentrations attained with the *DDD–AAA* system than with the *DAD-ADA* H-bonding pattern, as a result of a K_{ref} that is about 4 orders of magnitude higher.

3.2. Conclusions.

We would like to underline that, aside from their exceptional thermodynamic stability, the *c***G1-B1-C1**$_4$/ *c***iG1-B1-iC1**$_4$ / *c***A1-B1-U1**$_4$ macrocycles studied in this Chapter constitute a kinetically stabilized species in the overall self-assembly landscape. No matter the concentration, temperature or solvent conditions employed, the cyclic tetramers are always observed as a slowly exchanging species in the NMR timescale, exhibiting exchange rate constants within the 1-10 s^{-1} range. Furthermore, the H-bonding pattern is so important, and clearly demonstrate that cyclic systems constructed from symmetric *DAD-ADA* H-bonding pairs are much less stable than the homologues assembled from unsymmetric *ADD-DAA* or *DDA-AAD* pairs. On one hand, the *DAD-ADA* bonding pattern reduces considerably the enthalpy of intermolecular association due to the absence of attractive secondary interactions between vicinal H-bonding groups. On the other, the symmetry of this pattern introduces the possibility of multiple binding modes and hence a higher number of degrees of freedom in linear oligomers, which are then lost upon macrocyclization. This effect, of entropic origin, has a large impact on the *EM* of the system, which in our case is reduced by about 3 orders of magnitude.

Our conclusions could in principle be extended to many linear or cyclic supramolecular systems assembled *via* multipoint binding interactions. If a discrete, well-defined closed architecture is to be designed, rigid monomers with a suitable geometry in combination with an unsymmetric binding motif should be used to enhance the *EM*s of the cyclic system(s). In other words, the binding interaction should also contribute to the preorganization of the system towards a specific cycle, reducing the degrees of freedom of any other competitive supramolecular species. If, on the other hand, linear supramolecular polymers are pursued, a symmetric multipoint binding interaction would be the best choice to minimize chelate cooperativity, and hence the tendency of the supramolecular system to form undesired cycles.

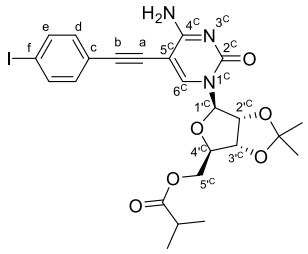
3.3 Experimental Section.

The General Methods detailed in the Experimental Section **1.4** of *Chapter 1* and Experimental Section **2.3** of *Chapter 2* are also applicable here. The work described in this Chapter can also be found in the Supporting Information of our papers: *Angew. Chem. Int. Ed.* **2015**, *54*, 6780–6784 (VIP Paper) (Ref: 156) and *Angew. Chem. Int. Ed.* **2016**, *55*, 223–227 (Ref: 158).

3.3.1 Synthesis and Characterization.

Monocoupling products

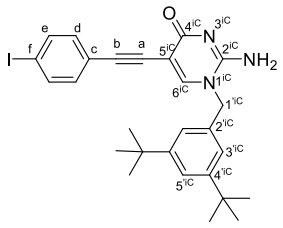
Cytidine monocoupling products.



C1-B1-I. C1-B1-I was prepared according to a *Standard Procedure C* between the ethynyl-nucleobase **C1** and 1,4-diiodobenzene. A dry a THF/NEt_3 (4:1) mixture (8 mL) was poured over **C1** (566 mg, 1.5 mmol), 1,4-diiodobenzene (3.5 g, 10.5 mmol), $PdCl_2(PPh_3)_2$ (9.5 mg, 0.01 mmol), and CuI (1.3 mg, 7 µmol). The mixture was stirred under argon for 12 h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording **C1-B1-I** as a yellow solid (730 mg, 84 %). The excess of 1,4-diiodobenzene was recovered.

^{1}H RMN (300 MHz, $CDCl_3$) δ(ppm) = 8.57 (s, 1H, NH^{4C}), 7.67 (d, J = 4.1 Hz, 2H, H^d), 7.63 (s, 1H,H^{6C}), 7.12 (d, J = 8.4 Hz, 2H, H^e), 5.83 (s, 1H, NH^{4C}), 5.65 (d, J = 1.7 Hz, 1H, $H^{1'C}$), 4.90 (dd, J = 6.3, J' = 1.7 Hz, 1H, $H^{2'C}$), 4.73 (dd, J = 6.3, J' = 3.7 Hz, 1H, $H^{3'C}$), 4.47 – 4.14 (m, 3H, $H^{4'C}$, $H^{5'C}$), 2.50 (hep, 1H, J = 7.0 Hz, -OCOCH(CH_3)$_2$), 1.51 (s, 3H, -OC(CH_3)), 1.28 (s, 3H, -OC(CH_3)), 1.08 (dd, J = 7.0, J' = 2.5 Hz, 6H, -COCH-(CH_3)$_2$).

*Iso*cytidine monocoupling product.



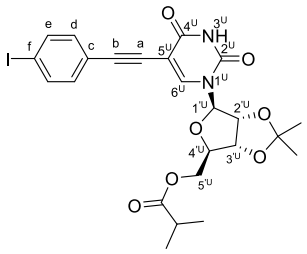
iC1-B1-I. iC1-B1-I was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between the ethynyl-nucleobase **iC1** and **B1**. A dry a THF/NEt_3 (4:1) mixture (40 mL) was poured over **B1** (3.6 g 11.1 mmol), $PdCl_2(PPh_3)_2$ (15 mg, 0.02 mmol), and CuI (2.1 mg, 0.01 mmol). Then, **iC1** (0.4 g, 1.1 mmol) was added dropwise and the mixture was stirred under argon for 12h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (40:1), affording **iC1-B1-I** as an orange solid (0.49 g, 83 %). The excess of **B1** was recovered.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.58 (d, J = 8.3 Hz, 2H, H^e), 7.40 (s, 1H, H^{6iC}), 7.35 (s, 1H, $H^{5'iC}$), 7.14 (d, J = 8.3 Hz, 2H, H^d), 7.04 (s, 2H, $H^{3'iC}$), 4.98 (s, 2H, $H^{1'iC}$), 1.27 (s, 18H, -C(CH_3)$_3$).

^{13}C NMR (126 MHz, DMSO-d_6) δ(ppm) = 137.6, 132.8, 79.2, 79.2, 79.0, 78.7, 57.5, 45.8, 39.5, 34.5, 31.2, 23.1, 19.2, 13.5.

MS (FAB+): 540.1 $[M+H]^+$.

Uridine monocoupling product.



U1-B1-I. U1-B1-I was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between the ethynyl-nucleobase **U1** and **B1**. A dry a THF/NEt_3 (4:1) mixture (50 mL) was poured over **B1** (4.4 g 13.2 mmol), $PdCl_2(PPh_3)_2$ (37 mg, 0.05 mmol), and CuI (5.0 mg, 0.03 mmol). Then, **U1** (1.0 g, 2.6 mmol) was added dropwise and the mixture was stirred under argon for 12h at 40 ºC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording **U1-B1-I** as a yellow solid (1.3 g, 85%). The excess of **B1** was recovered.

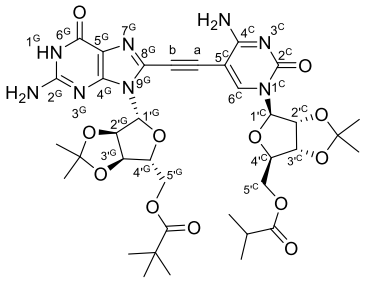
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.83 (s, 1H, N*H*), 8.12 (s, 1H, H^{6U}), 7.79 (d, J = 8.4 Hz, 2H, H^{e}), 7.24 (d, J = 8.4 Hz, 2H, H^{d}), 5.85 (d, J = 2.0 Hz, 1H, $H^{1'U}$), 5.07 (dd, J = 6.4, J' = 2.1 Hz, 1H, $H^{2'U}$), 4.81 (d, J = 6.7 Hz, 1H, $H^{3'U}$), 4.26 – 4.22 (m, J = 2.1 Hz, 3H, $H^{4'U}$, $H^{5'U}$), 2.58 (hept, J = 7.0 Hz, 1H, -OCOC*H*$(CH_3)_2$), 1.49 (s, 3H, -OC(CH_3)), 1.30 (s, 3H, -OC(CH_3)), 1.07 (dd, J = 7.0, J' = 4.2 Hz, 6H, -OCOCH(C$H_3)_2$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 176.6, 161.3, 149.1, 143.7, 137.6, 133.1, 122.0, 114.9, 100.4, 94.8, 94.0, 93.1, 85.1, 84.9, 81.3, 80.8, 63.8, 34.0, 27.2, 25.4, 19.1, 19.0.

MS (FAB+): 581.2 $[M+H]^+$.

Monomers

Synthesis of the monomers toward one-component tetramers.

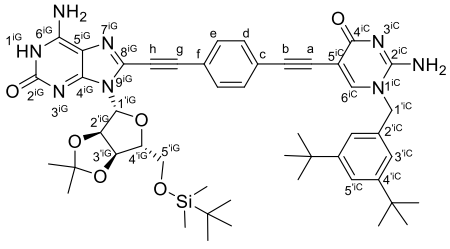


G1-C1. G1-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between iodo-nucleobase **C1.1** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (10 mL) was poured over **G1** (200 mg 0.5 mmol), $PdCl_2(PPh_3)_2$ (6.5 mg, 9.3 µmol), and CuI (0.88 mg, 4.6 µmol). Then, **C1.1** (289 mg, 0.6 mmol) was added and the mixture was stirred under argon for 12h at 40 ºC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by recrystallization in acetonitrile, affording **G1-C1** as a white solid (316 mg, 81 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.00 (s (broad), 1H, NH^{1G}), 8.28 (s, 1H, H^{6C}), 8.00 (s (broad), 1H, NH^{4C}), 7.37 (s (broad), 1H, NH^{4C}), 6.80 (s (broad), 2H, NH_2^{2G}), 6.13 (s, 1H, $H^{1'G}$), 5.81 (s, 1H, $H^{1'C}$), 5.43 (d, J = 6.3 Hz, 1H, $H^{2'G}$), 5.22 (dd, J = 6.2, J' = 4.0 Hz, 1H, $H^{2'C}$), 5.02 (dd, J = 6.4, J' = 1.6 Hz, 1H, $H^{3'G}$), 4.91 – 4.76 (m, 1H, $H^{3'C}$), 4.42 – 3.97 (m, 6H, $H^{4'G}$, $H^{5'G}$, $H^{4'C}$, $H^{5'C}$), 2.61 – 2.54 (m, 1H, COC*H*), 1.50 (d, J = 7.9 Hz, 6H, -OC(CH_3)), 1.31 (d, J = 9.8 Hz, 6H, -OC(CH_3)), 1.08 (d, J = 10.3 Hz, 14H, -COCH-(C$H_3)_2$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 179.4, 177.8, 164.6, 159.5, 155.7, 154.4, 150.8, 145.0, 121.8, 118.1, 114.2, 113.4, 95.8, 94.6, 93.2, 92.4, 89.3, 87.7, 85.4, 85.0, 84.2, 83.5, 81.4, 79.4, 79.1, 77.8, 76.5, 76.4, 76.1, 66.3, 64.2, 38.1, 33.2, 29.4, 27.8, 27.6, 27.5, 25.8, 25.6, 19.5.

HRMS (FAB+): Calculated for $C_{36}H_{47}N_8O_{12}$: 783.3313 $[M+H]^+$. Found: 783.3334 $[M+H]^+$**.**



iG1-B1-iC1. iG1-B1-iC1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between **iC1-B1-I** and the ethynyl-nucleobase **iG1**. A dry a THF/NEt_3 (4:1) mixture (10 mL) was poured over **iC1-B1-I** (88.3 mg, 0.2 mmol), $PdCl_2(PPh_3)_2$ (2.3 mg, 3 µmol), and CuI (0.3 mg, 1 µmol). Then, **iG1** (90 mg, 0.2 mmol), was added dropwise and the mixture was stirred under argon for 12h at 40 ºC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by recrystallization in acetonitrile, affording **iG-B1-iC** as an orange solid (129.6 mg, 91 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.16 (s (broad), 1H, NH^{1iG}), 8.40 (s, 1H, H^{6iC}), 8.20 (s, 1H, $H^{5'iC}$), 7.75 (d, J = 8.0 Hz, 2H, H^{e}), 7.66 (d, J = 8.2 Hz, 2H, H^{d}), 7.49 (s, 2H, NH^2, $H^{3'iC}$), 6.21 (s, 1H, NH^1), 5.75 (d, J =

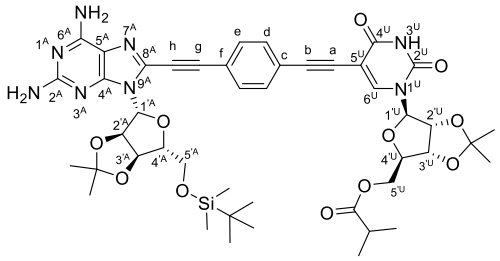
6.1 Hz, 1H, $H^{1'iG}$), 5.20 (s, 2H, $H^{1'iC}$), 5.13 (s, 1H, $H^{2'iG}$), 4.26 (s, 1H, $H^{3'iG}$), 3.96 – 3.82 (m, 2H, $H^{5'iG}$), 3.54 (s, 1H, $H^{4'iG}$), 1.69 (s, 3H, -OC(CH_3)), 1.49 (s, 3H, -OC(CH_3)), 1.43 (s, 18H, -C(CH_3)$_3$), 0.94 (s, 9H, -Si(CH$_3$)$_2$(CH_3)$_3$), 0.07 (s, 6H, -Si(CH_3)$_2$(CH$_3$)$_3$).

^{13}C NMR (126 MHz, DMSO-d_6) δ(ppm) = 167.5, 154.6, 150.8, 147.5, 134.6, 131.9, 131.5, 130.5, 128.8, 128.7, 124.4, 121.5, 121.1, 119.6, 113.2, 102.2, 93.2, 90.9, 89.1, 87.3, 86.2, 82.3, 81.8, 81.5, 80.4, 79.2, 79.0, 78.7, 63.3, 61.9, 53.4, 39.5, 34.6, 31.2, 29.0, 27.4, 27.1, 25.7, 25.3, 23.1, 22.1, 19.2, 18.0, 14.0, 13.5, -3.2, -5.4, -5.5.

HRMS (FAB+): Calculated for $C_{48}H_{66}N_8O_6Si$: 873.4405 $[M+H]^+$. Found: 873.4496 $[M+H]^+$.

UV-VIS: λ_{max} (monomer, DMAC, 298 K) = 371 nm (ε = 36152 $M^{-1}cm^{-1}$).

Emission: λ_{max} (monomer, DMAC, 298 K) = 463 nm.



A1-B1-U1. A1-B1-U1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between **U1-B1-I** and the ethynyl-nucleobase **A1**. A dry a THF/NEt$_3$ (4:1) mixture (10 mL) was poured over **U1-B1-I** (0.5 g, 0.9 mmol), PdCl$_2$(PPh$_3$)$_2$ (12 mg, 17 µmol), and CuI (1.6 mg, 9 µmol). Then, **A1** (0.6 g, 1 mmol), was added dropwise and the mixture was stirred under argon for 12h at 40 ^{o}C. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by recrystallization in acetonitrile, affording **A1-B1-U1** as a yellow solid (129.6 mg, 91 %). The product was purified by column chromatography on silica gel eluted with CHCl$_3$/MeOH (40:1), affording **A1-B1-U1** as a yellow solid (0.5 g, 62 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.87 (s (broad), 1H, NH^{3U}), 8.16 (s, 1H, H^{6U}), 7.65 (d, J = 8.3 Hz, 2H, H^e), 7.55 (d, J = 8.3 Hz, 2H, H^d), 7.01 (s (broad), 2H, NH_2^{2A}), 6.21 (s (broad), 2H, NH_2^{6A}), 6.12 (d, J = 1.6 Hz, 1H, $H^{1'A}$), 5.86 (d, J = 2.0 Hz, 1H, $H^{1'U}$), 5.60 – 5.55 (m, 1H, $H^{2'A}$), 5.14 (dd, J = 6.2, J' = 3.4 Hz, 1H, $H^{2'U}$), 5.08 (dd, J = 6.4, J' = 2.1 Hz, 1H, $H^{3'A}$), 4.82 (d, J = 6.5 Hz, 1H, $H^{3'U}$), 4.26 (s, 3H, $H^{4'A}$, $H^{5'A}$), 4.11 (d, J = 3.9 Hz, 1H, $H^{4'U}$), 3.76 – 3.65 (m, 2H, $H^{5'U}$), 2.58 (hept, J = 7.0 Hz, 1H, -OCO-CH(CH$_3$)$_2$), 1.54 (s, 3H, -OC(CH_3)), 1.50 (s, 3H, -OC(CH_3)), 1.35 (s, 3H, -OC(CH_3)), 1.30 (s, 3H, -OC(CH_3)), 1.08 (dd, J = 7.0, J' = 4.0 Hz, 6H, -OCOCH-(CH_3)$_2$), 0.76 (s, 9H, -OSi(CH$_3$)$_2$-(CH_3)$_3$), -0.15 (d, J = 5.6 Hz, 6H, -OSi(CH_3)$_2$-(CH$_3$)$_3$).

^{13}C NMR (75 MHz, CDCl$_3$) δ(ppm) = 176.9, 162.5, 160.5, 157.3, 150.9, 150.3, 143.1, 132.2, 131.8, 131.6, 122.9, 121.5, 115.8, 115.4, 113.6, 100.7, 94.0, 90.6, 87.5, 83.3, 82.4, 81.5, 81.0, 80.1, 77.2, 63.5, 33.9, 32.1, 29.8, 29.5, 27.6, 27.2, 26.7, 26.0, 25.7, 22.8, 19.4, 19.0, 18.5, 14.3, 1.2, -5.2, -5.3.

HRMS (MALDI): Calculated for $C_{45}H_{56}N_8O_{11}Si$: 913.3838 $[M+H]^+$. Found: 935.3730 $[M+Na]^+$. Matrix: DCTB

UV-VIS: λ_{max} (monomer, Dioxane, 298 K) = 365 nm (ε = 38767 $M^{-1}cm^{-1}$).

Emission: λ_{max} (monomer, Dioxane, 298 K) = 443 nm.

3.3.2 NMR and Optical Spectroscopy Dilution and Titration Experiments.

NMR dilutions and titrations were carried out in in 5 mm NMR tubes using DMSO-d_6, DMF-d_7, THF-d_8, acetone-d_6, CDCl$_2$CDCl$_2$, CDCl$_3$ or 2:3 *v/v* CHCl$_3$-CCl$_4$ as solvents. Deuterated solvents were purchased from Aldrich in ampoules and used as received. Residual CHCl$_3$ was used as the internal references (7.26 ppm), respectively. UV-vis dilutions and titrations were carried out in THF, CHCl$_3$ or 2:3 *v/v* CHCl$_3$-CCl$_4$ (Alfa Aesar, Spectrophotometric Grade). The experiments were performed in 1 cm or 1 mm path length quartz cuvettes. Volumes were added using Hamilton microsyringes. UV-vis absorbances were kept within the 0.2-3.5 range. Temperature control was set at 298 K in all cases.

Dilution experiments were carried out by successive injections of a stock solution of the corresponding dinucleoside monomer into clean solvent, thus increasing the concentration along the experiment. We found this method more practical and reliable than performing successive dilutions of the concentrated starting sample. The full ^{1}H NMR/UV-vis spectra were recorded over at least 15 concentrations, considering, as far as possible, that most of them

should yield chemical shift/absorbance data within the 20-80% saturation range. Hence, the concentration range targeted depended on the dimerization constant expected for each nucleobase. Each dilution experiment was repeated at least twice.

Titration experiments were performed as follows. A sample of the host nucleoside was dissolved in the appropriate solvent, whose concentration, indicated in each experiment below, varied depending on the technique employed (^{1}H NMR or UV-Vis) and the expected magnitude of the association constant. A portion of this solution was used as the host sample, and the remainder was used to dissolve the sample of the guest, so that the host concentration remained constant throughout the titration. Successive aliquots of the guest solution, typically more concentrated, were added to the host sample, and the whole ^{1}H NMR / UV-vis spectra were recorded after each of the 15-20 guest additions. Again, in order to cover as much as possible the 20%-80% probability of binding range, the initial host concentration and the number of guest equivalents targeted was lower or higher as a function of the expected association constant between complementary bases. Each titration experiment was repeated at least twice.

The treatment necessary to calculate K_T, K_C, ΔH and ΔS through optical spectroscopy techniques is also commented in the Supporting Information of our papers.[156,158]

Chapter 4

Self-Sorting Governed by Chelate Cooperativity

Self-sorting phenomena has been studied in *Chapter 4*, in which the narcissistic macrocyclization behaviour of our three dinucleoside monomers has been inquired by comparing the self-assembly in solution of their binary mixtures with quaternary mixtures of the corresponding mononucleosides. The self-assembly of these mixed systems has been characterized either by ^{1}H and NOESY NMR or by CD and fluorescence spectroscopy using mono- and dinucleosides featuring a resonance energy donor or acceptor dyes as central blocks. The NMR results are displayed in this Chapter as part of the work developed during this Thesis, whereas only the most relevant preliminary results with the FRET dyes are included here in the final Section, since they were carried out in parallel by Jorge Camacho and then continued by David Serrano, both PhD students in the *MSMn* group.

Most of the work presented in *Chapter 4* will be published in the following article, which is under preparation:

"Self-sorting Phenomena Governed by Chelate Cooperativity", C. Montoro-García, D. Serrano-Molina, M. J. Mayoral, D. González-Rodríguez.

4.1 Strategy Toward Self-Sorting Phenomena.

As shown in Figure 4.1a, **G** and **C** bases display an unsymmetric *ADD-DAA* H-bonding pattern through their Watson-Crick edges that is not complementary to the symmetric *DAD-ADA* pattern shown by **A** and **U**, but that can be complementary to the *DDA-AAD* pattern displayed by **iG** and **iC** when considering a reverse Watson-Crick binding configuration. This means that **G** can bind to **iC** and **iG** to **C** with similar strength as in the regular Watson-Crick **G**:**C** and **iG**:**iC** configuration. As a matter of fact, the association constants of each possible pair between these four bases were calculated in $CDCl_3$ in *Chapter 2* as: K_a (**G**:**C**) = 1.5–3.0 x 10^4 M^{-1}, K_a (**iG**:**iC**) = 2.2–4.7 x 10^4 M^{-1}, K_a (**G**:**iC**) = 1.2–2.0 x 10^4 M^{-1}, and K_a (**iG**:**C**) = 2.0–4.3 x 10^4 M^{-1}. Figure 4.1b schematically represents self-sorting phenomena when mono- and dinucleosides are mixed separately.

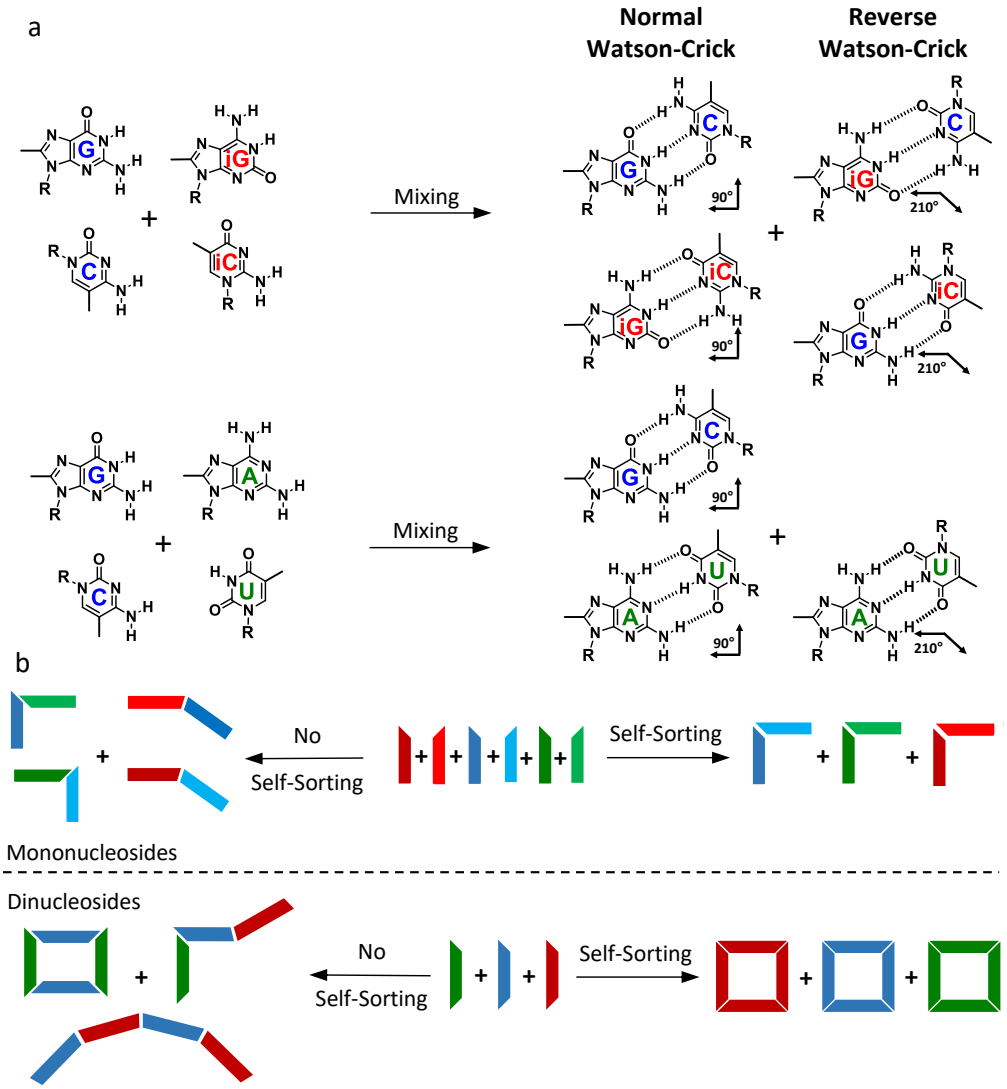


Figure 4.1. Different Watson-Crick H-bonded base-pairs studied in this *Chapter 4*. (**a**) Possible normal and reverse Watson Crick Watson-Crick H-bonded base-pairs and (**b**) scheme of the possible self-sorting or not self-sorting phenomena.

The set of mononucleoside and dinucleoside molecules employed in this study is shown schematically in Figure 4.2. Two main modifications were included in the general structure of these monomers. The first one is the nature of the base in the mononucleoside (G, C, A, U, iG iC) or the complementary base pair in the dinucleoside (G-C, A-U, iG-iC). The second one is the kind of central block attached to these bases, which can be either a *p*-phenylene ring or a linearly disubstituted dye (bithiophene or BODIPY). These dyes were selected taking into account: 1) their identical length, so that the formation of mixed cyclic assemblies remains possible and self-sorting is not driven by size effects and, 2) their absorption and emission features, so that they constitute a threesome of FRET donors and acceptors. Within the bithiophene-BODIPY1 and BODIPY-BODIPY2, primary donor and secondary acceptor. Molecules featuring a central *p*-phenylene ring were studied by NMR whereas the self-

assembly of the dye monomers was examined by circular dichroism (CD) and fluorescence spectroscopy.

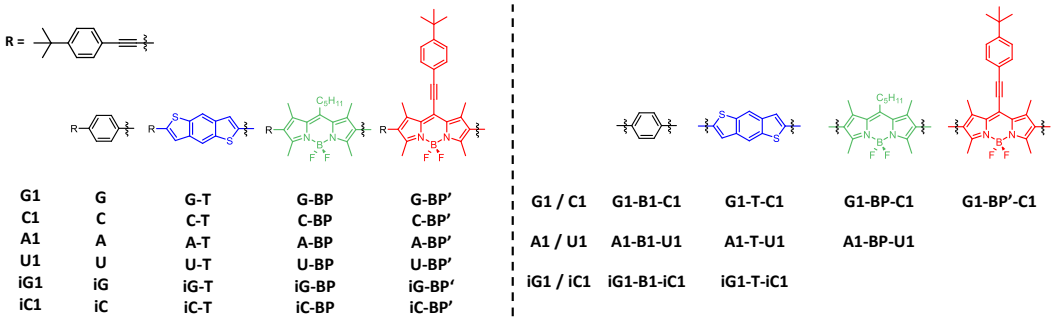


Figure 4.2. Chemical structure of the mono- and dinucleoside derivatives endowed with different central blocks employed in this *Chapter4*.

4.1.1 Quaternary Mononucleoside and Binary Dinucleoside Combinations studied by NMR.

The focus of this *Chapter 4* is hence to study the supramolecular behavior of mixtures of dinucleosides, which are able to assemble in cyclic tetramer species,[158] and compare them with the corresponding mixtures of mononucleosides, which are expected to self-associate in pairs through their Watson-Crick edges. Therefore, we started by examining the ^{1}H NMR spectra in 1:1 mixtures of complementary mononucleosides (**G**+**C**, **A**+**U** and **iG**+**iC**) at a fixed concentration (10^{-2} M) and temperature (298 K) in $CDCl_3$. As can be observed in Figure 4.3a-b, H-bonding formation between complementary pairs becomes evident when examining the upfield shift experienced by the different protons involved. These mononucleoside pairs were further combined in 1:1:1:1 mixtures (**G**+**C**+**A**+**U** and **G**+**C**+**iG**+**iC**). In the case of the **G**+**C**+**A**+**U** mixture no significant changes in the ^{1}H NMR spectrum were detected. However, for the **G**+**C**+**iG**+**iC** combination some minor shifts of some signals were observed upon mixing.

2D NOESY experiments performed in the same conditions could confirm the proximity of the relevant H-bonded protons in the complementary pairs and provide an assessment whether they self-sort or not in their 1:1:1:1 mixtures. As shown in Figures 4.3c, the **G**+**C**+**iG**+**iC** mixture exhibit cross-peaks between all possible combinations of Watson Crick and reverse Watson-Crick pairs (**G**:**C**, **iG**:**iC**, **G**:**iC** and **iG**:**C**), but also between **G** and **iG**. To our surprise, the **G**+**C**+**A**+**U** mixture also displayed cross-peaks between all possible pairs (**G**-**C**, **A**-**U**, **G**-**U**, **A**-**C**, **G**-**A** and **C**-**U**). This may be due to the formation of non-complementary pairs and/or to the association in higher-order species, but in any case these results clearly show that no binding selectivity is observed in the quaternary mononucleoside mixtures and any kind of self-sorting phenomena is totally absent.

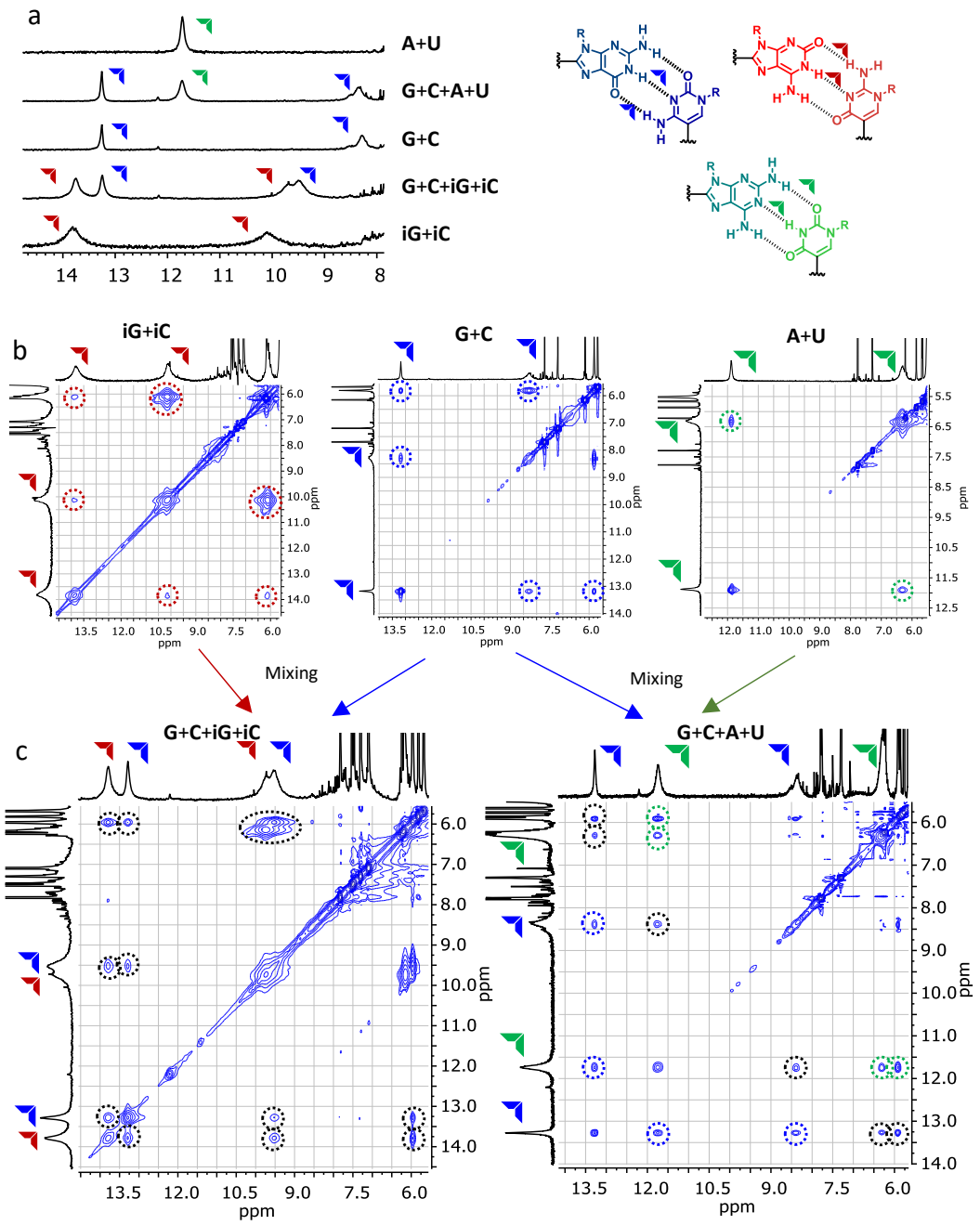


Figure 4.3. *Quaternary mononucleoside combinations examined by ^{1}H and 2D NOESY NMR.* (**a**) Downfield region of the ^{1}H-NMR spectra showing the chemical shift of the complementary Watson-Crick pairs (**G+C**, **iG+iC**, **A+U**) and their mixtures (1:1:1:1 mixture of **G+C+iG+iC** and **G+C+A+U** respectively). Region of the NOESY NMR spectrum in $CDCl_3$ of (**b**) 1:1 complex of **G+C**, **iG+iC**, **A+U** respectively, (**c**) 1:1:1:1 mixture of **G+C+iG+iC** and **G+C+A+U** respectively. C = 1.0 x 10^{-2} M in $CDCl_3$ and T = 298 K in all cases.

We then turned our attention to the behavior of 1:1 mixtures of dinucleosides in identical conditions. As shown in Figure 4.4a, mixing the non-complementary **G1-B1-C1**+**A1-B1-U1** dinucleosides produced no change in their ^{1}H NMR spectra. NOESY experiments furthermore reveals that G only binds to C, while A only binds to U. Therefore, the **G1-B1-C1**+ **A1-B1-U1** mixture exhibits clear narcissistic self-sorting characteristics, which is what we would expect in view of the non-complementary H-bonding pattern of the 2 Watson-Crick pairs involved. Now, in the case of the **G1-B1-C1**+**iG1-B1-iC1** combination the H-bonding patterns are complementary and, as discussed before, these 4 bases could in principle bind through all combinations of Watson Crick and reverse Watson-Crick pairs, which would lead to a complex mixture of cyclic and open oligomers. However, ^{1}H NMR and NOESY spectra clearly show that only their respective cyclic tetramers are formed, where G only binds to C whereas iG binds exclusively to iC, and G-iC or iG-C cross-peaks were not detected. Therefore, in this particular case, *narcissistic self-sorting is clearly not ruled by H-bonding complementarity, but by chelate cooperativity*, that is, by the strong tendency of both dinucleoside molecules to form cyclic tetramers with high *EM*s. Only when **G1-B1-C1** and **iG1-B1-iC1** associate independently, the cyclic tetramer species can be assembled because a Watson-Crick 90° angle is required.

Unfortunately, we could not be able to study the 1:1 mixtures of **iG1-B1-iC1** + **A1-B1-U1** due to solubility and stability problems. On the one hand, Figure 4.5a shows clearly the solubility problems displayed by **G1-B1-C1** + **A1-B1-U1** in CCl_4. As increasing amounts of $CDCl_3$ content were added, the solubility of both monomers is enhanced and *c***A1-B1-U1**$_4$ could be formed quantitatively in a $CDCl_3$:CCl_4 (2:3) mixture. In pure $CDCl_3$, the *c***A1-B1-U1**$_4$ assembly is not formed quantitatively, as explained in *Chapter 3*. On the other, Figure 4.5b reveals the solubility problems of **iG1-B1-iC1** in $CDCl_3$ which can be solved by the addition of 1% DMSO or directly using a more polar solvent such as pure THF-d_8. Both problems preclude the measurement of 1:1 mixtures of **iG1-B1-iC1**+**A1-B1-U1**. In the first scenario, **iG1-B1-iC1** is completely insoluble and, in the second scenario, **A1-B1-U1** is not be able to form discrete cyclic systems with a few amounts of DMSO or in pure THF and the monomer is the most abundant entity present.

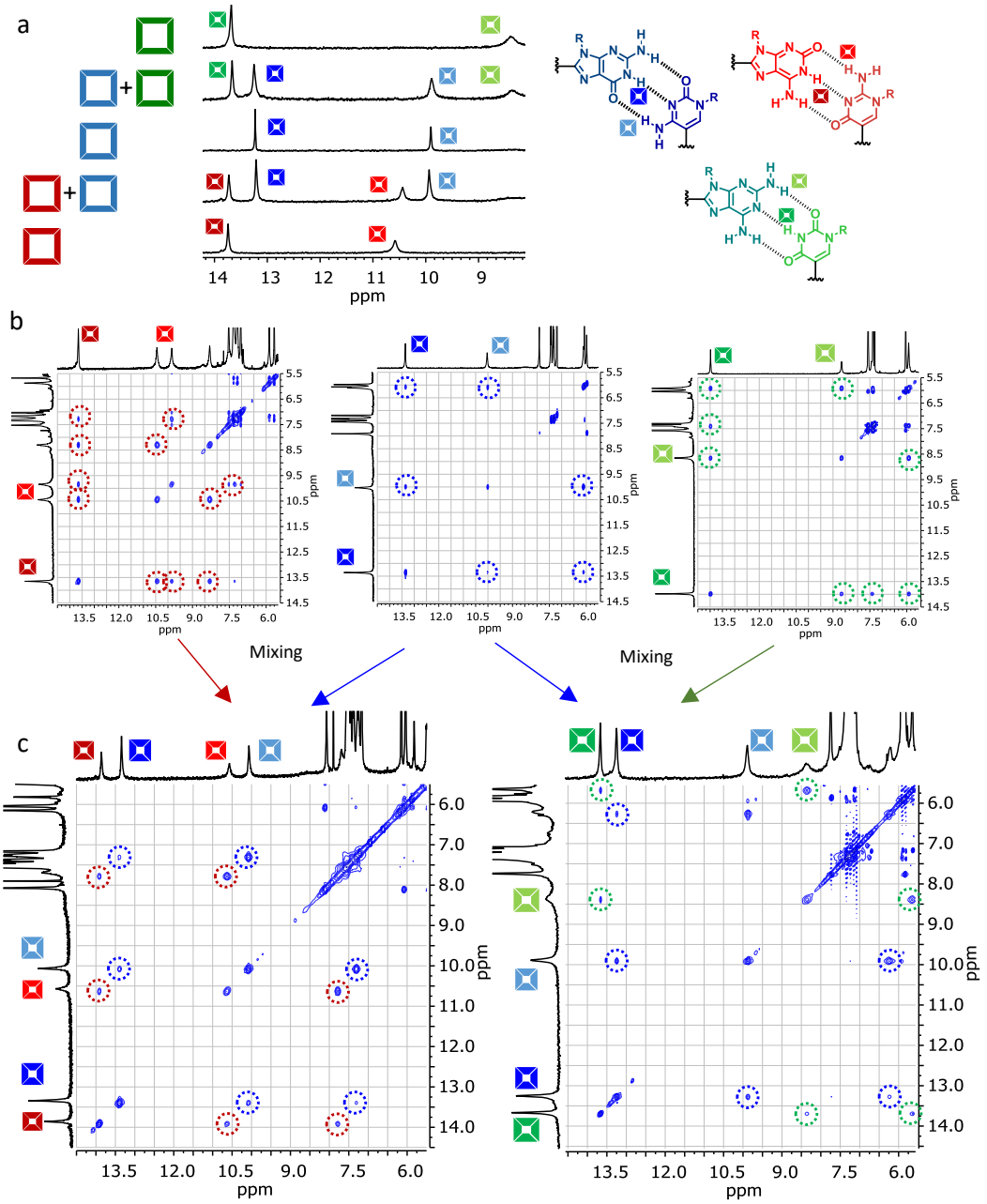


Figure 4.4. *Binary dinucleoside combinations examined by ^{1}H and 2D NOESY NMR. Cyclic tetramer self-sorting.* Downfield region of the (**a**) ^{1}H-NMR spectra and (**b**) of the NOESY NMR spectra of *c***G1-B1-C1**$_4$ (THF), *c***iG1-B1-iC1**$_4$ (THF) and *c***A1-B1-U1**$_4$, and their respective 1:1 mixtures of *c***A1-B1-U1**$_4$ + *c***G1-B1-C1**$_4$ ($CDCl_3$:CCl_4 / 2:3) and *c***G1-B1-C1**$_4$ + *c***iG1-B1-iC1**$_4$ (THF) showing cross-peaks between the H-bonded proton signals. T = 298 K and C = 1.0 x 10^{-2} M in all cases.

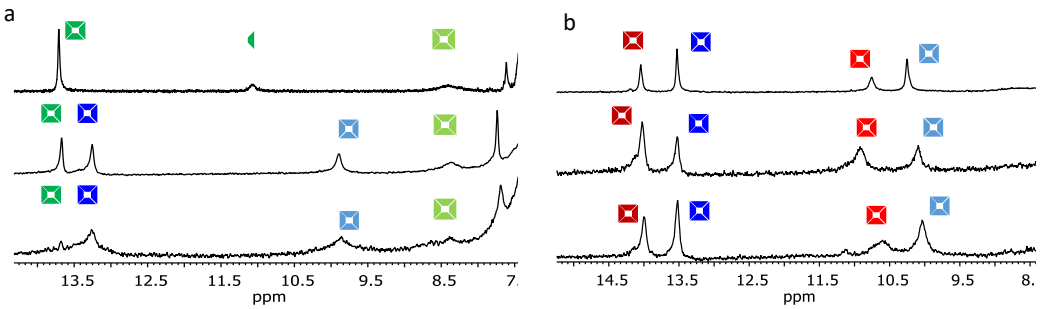


Figure 4.5 *Tuning solvent composition for **A1-B1-U1** and **iG1-B1-iC1***. Downfield region of the ^{1}H NMR spectra of (**a**) **G1-B1-C1** + **A1-B1-U1** in CCl_4 with increasing amounts of $CDCl_3$, (**b**) Temperature-dependent measurements of **G1-B1-C1** + **A1-B1-U1** in $CDCl_3$, and (**d**) **G1-B1-C1** + **iG1-B1-iC1** in $CDCl_3$ with increasing amounts of DMSO and finally in THF.

It should be noted that the absence of self-sorting phenomena would not only lead to open oligomeric species. Other cyclic tetramers could be formed by double H-bond A:C or G:U association, as shown in Figure 4.6. In addition, if the 210°reverse Watson-Crick G-iC or iG:C association is established, an unstrained cyclic dodecamer may be formed in solution as well, as shown in Figure 4.7. However, it is clear from the NOESY experiments that these mixed cyclic assemblies are not formed. This is, on the other hand, quite logical, since in the first case only two H-bonds are formed and the binding constants should be lower, while in the second, the *EM* of the hypothetic dodecamer should be very low, certainly much lower than those observed for our cyclic tetramers.

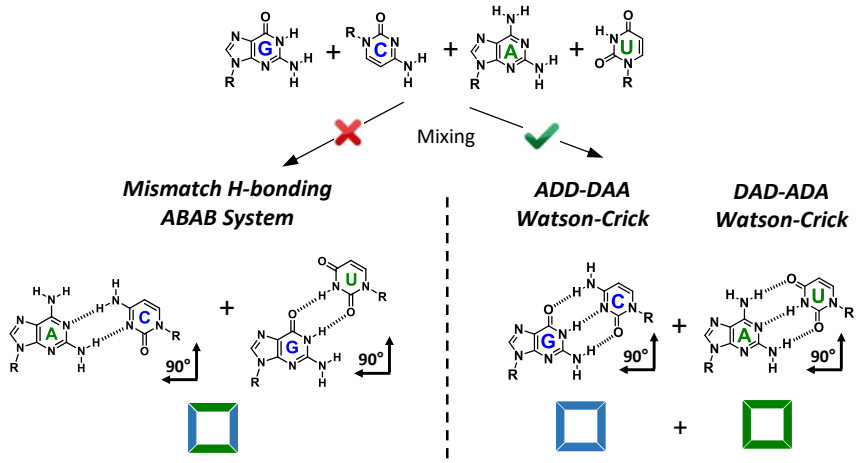


Figure 4.6. **Unsymmetric versus Symmetric H-bonding patterns**: It is possible to form a ABAB cyclic tetramer system through the mismatch H-bonding interaction between **G**:**U** and **A**:**C**. However, the only two species presents in the ^{1}H NMR spectra are the tetramers formed by complementary Watson-Crick H-bonding between **G**:**C** and **A**:**U** due to it is one-component system and K_{ass} is quite high in comparison with the previous one.

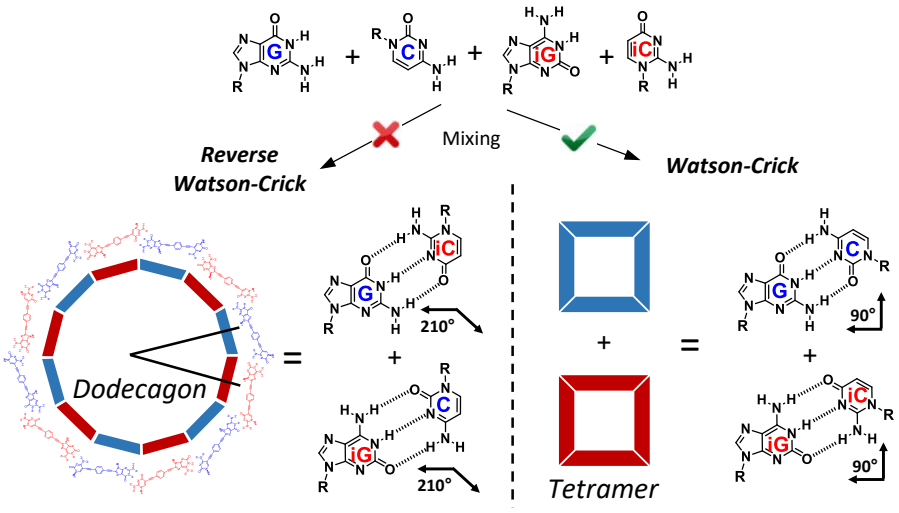


Figure 4.7. **Symmetric H-bonding patterns between nucleobases**: Due to the symmetric H-bonding patterns shared by **G**, **C**, **iG**, **iC**, they can self-assemble in Watson-Crick arrangement leading to ring-closure (**G**:**C** and **iG**:**iC**) or in reverse Watson-Crick arrangement in which a AB dodecagon system could be formed (**G**:**iC** + **iG**:**C**). Dodecagon cannot be formed because the theory predicts that the less strained and the less size macrocycle will be the most stable entity in solution. Hence, the *EM* governs this macrocyclization.

Cyclic Tetramer Dissociation.

In view of the different stability observed between A-U and G-C/iG-iC cyclic tetramers, we thought it would be possible to progressively and selectively dissociate the weaker macrocycle in the presence of the stronger one.

For instance, increasing temperature in $CDCl_3$ solutions of the **G1-B1-C1**+**A1-B1-U1** mixture within the 253-323 K range (see Figure 4.8a) resulted in the gradual dissociation of only the *c***A1-B1-U1**$_4$ macrocycle, whereas *c***G1-B1-C1**$_4$ remains intact in the whole temperature range. On the other hand, addition of DMSO-d_6 to (2:3) $CDCl_3$:CCl_4 solutions[164] of **G1-B1-C1**+**A1-B1-U1** led to the observation of two clear regimes (Figure 4.8b). In the first one, from 0 to 12% v/v of DMSO-d_6, *c***A1-B1-U1**$_4$ is progressively dissociated in the presence of the stronger *c***G1-B1-C1**$_4$ macrocyle, which show no sign of denaturation. This is evidenced by the appearance of the **A1-B1-U1** monomer U-imide signal at *ca.* 11.8 ppm. In the second regime, starting over *ca.* 20% DMSO-d_6, *c***G1-B1-C1**$_4$ is then dissociated to the monomeric species, showing the **G1-B1-C1** G-amide signal at 10.9 ppm. It should be remarked that both cyclic tetramers are in slow exchange in the NMR timescale with their respective monomeric species. Figure4.8c shows the same DMSO-d_6 titrations with the **G1-B1-C1**+**iG1-B1-iC1** mixture. In this case, due to the similar K_a and *EM* values displayed by *c***G1-B1-C1**$_4$ and *c***iG1-B1-iC1**$_4$,[158] cyclic tetramer dissociation occurs in parallel and the G/iG-amide signals of both **G1-B1-C1** and **iG1-B1-iC1** monomers are detected in slow exchange at *ca.* 10.8 ppm after a DMSO-d_6 volume fraction of 80%.

[164] A 2:3 $CDCl_3$:CCl_4 mixture had to be employed with **A1-B1-U1** instead of pure $CDCl_3$ in order to increase *c***A1-B1-U1**$_4$ stability and produce quantitative cyclic tetramer formation (see Figure 4.6 and please come back to *Chapter 3* for further details).

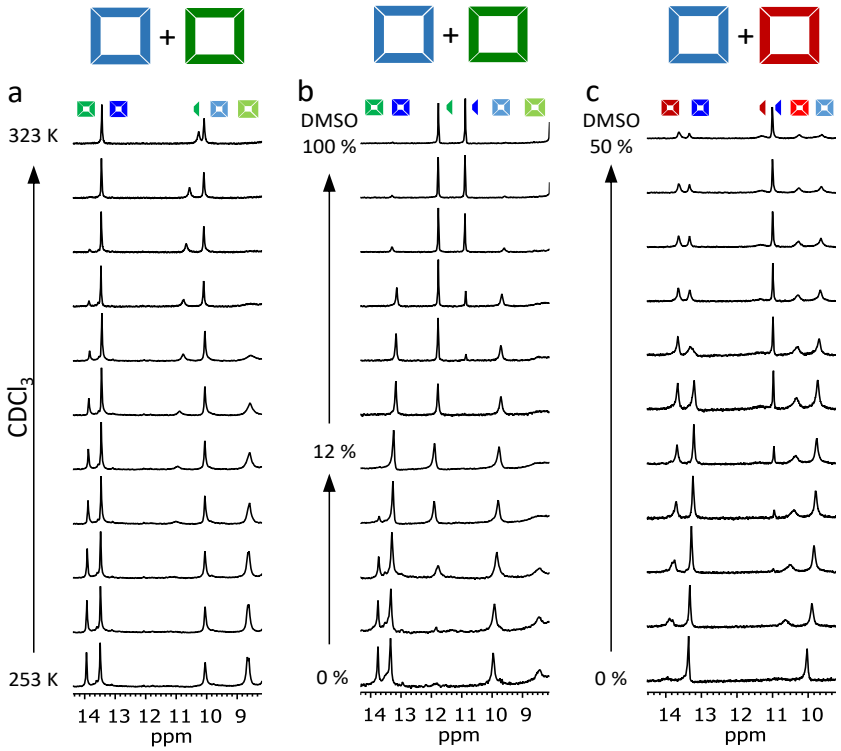


Figure 4.8 *Cyclic tetramer denaturation experiments.* Downfield region of the ^{1}H NMR spectra of (**a**) Temperature-dependent measurements of **G1-B1-C1** + **A1-B1-U1** in $CDCl_3$, (**b**) **G1-B1-C1** + **A1-B1-U1** in $CDCl_3$:CCl_4 (2:3) with increasing amounts of DMSO, and (**d**) **G1-B1-C1** + **iG1-B1-iC1** in $CDCl_3$ with increasing amounts of DMSO showing the slow exchange between tetramer-monomer proton signals.

4.1.2 Quaternary Mononucleoside and Binary/Ternary Dinucleoside Combinations studied by Optical Spectroscopy.

While NMR experiments already provided a reasonably clear picture of the self-assembly of mixtures of mono- and dinucleoside experiments, we complemented these studies with CD and fluorescence spectroscopy experiments using donor and acceptor FRET pairs (Figure 4.9a). As explained at the beginning of this chapter, only the most relevant preliminary results with the FRET dyes are shown and described below, which confirm our observations by NMR.

In these experiments, concentration was lowered to the 10^{-4}-10^{-6} M regime and toluene was used to increase binding strength between base pairs. As we determined in *Chapter 2*, association constants in this apolar solvent are increased in about one order of magnitude with respect to $CHCl_3$ (over 10^5 M^{-1} for G-C/iG-iC and over 10^3 M^{-1} for the A-U pair). We followed the same rationale as in the NMR experiments: the spectroscopic features of mononucleoside complementary pairs or of dinucleosides were examined first, and then the relevant 1:1 mixtures were generated and spectroscopic changes were monitored with time.

These results were then contrasted to the behavior of the 1:1 dinucleoside mixtures in the same conditions. We again first performed control experiments in which energy donor and acceptor couples were combined in monomers having the same base pairs: **GdC**+**GaC** and **AdU**+**AaU**, and recorded the spectroscopic changes experienced by the system as a function

of time (Figure 4.9b). At equilibrium, a statistical mixture of six different cyclic tetramers should be formed, since the central π–conjugated blocks have identical lengths and are end-capped with the same nucleobases. Donor and acceptor moieties are closely positioned in some of these macrocycles, thus allowing for resonance energy transfer to take place, which should be evidenced by a decrease of donor emission at the expense of an increase in acceptor emission. It is interesting to note that, in contrast to what was seen with the mononucleosides, equilibrium is reached very slowly with these dinucleoside mixtures in toluene, within a timescale of several hours, which underlines the extraordinarily high kinetic stability of the cyclic assemblies. When performing the same experiment with **GdC**+**GaC** in $CHCl_3$ or THF, equilibrium was instead reached within a few minutes.

We then studied the scenario where the bases in the dye monomers are different. **GdC**+**AaU** and **AdU**+**GaC** mixtures where examined first (Figure 4.9c). In sharp contrast to what was observed before, no change at all was detected over a period of 20 hours in the emission or CD spectra when these dinucleoside combinations were mixed together at either 10^{-4} M, 10^{-5} M or 10^{-6} M concentration in toluene or $CHCl_3$. The same results were found for the **iGdiC**+**AaU** mixture. This is in full agreement with the NMR results and confirms narcissistic self-sorting between cyclic tetramers when the bases are non-complementary in their H-bonding pattern. The question now arises whether a **iGdiC**+**GaC** mixture, having complementary pairs, would self-sort as well. As stated in Figure 4c when these dinucleosides are combined, their equilibrium mixture exhibits exactly the same spectroscopic features as the sum of the spectra when these samples are analyzed separately. This implies that **iGdiC** and **GaC** self-associate independently in their corresponding cyclic tetramers and no mixed assemblies, where G would bind to iC or iG to C, are formed. In short, these experiments using optical spectroscopy and dyes that absorb and emit in different spectral regions also support the notion that *narcissistic self-sorting is primarily governed by the strong chelate cooperativity manifested by each dinucleoside monomer when assembled as a cyclic tetramer*.

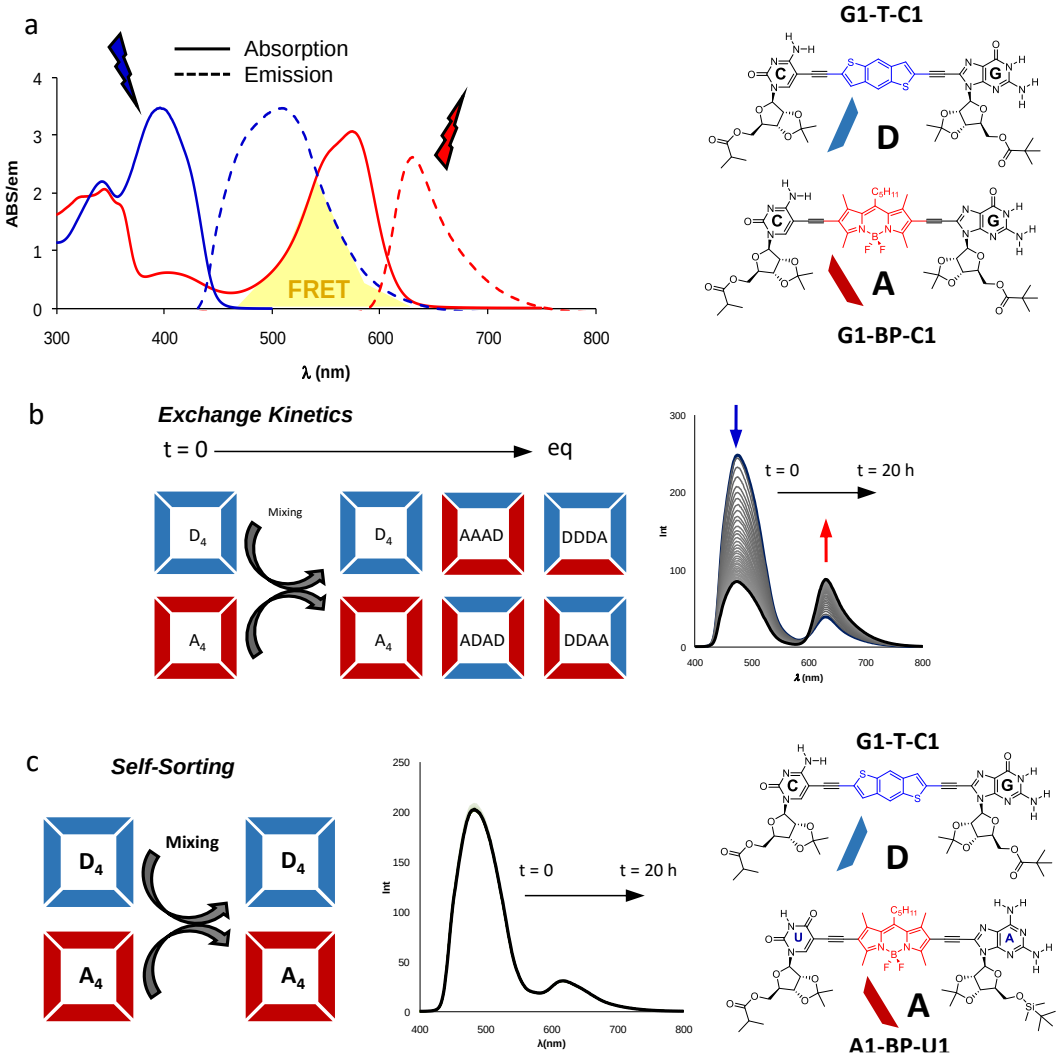


Figure 4.7 *Binary dinucleoside combinations examined by fluorescence spectroscopy. Self-sorting with donor-acceptor FRET dyes.* (**a**) Absorption and emission UV-vis spectra bands of the donor-acceptor FRET dyes measured separately, showing the FRET region. (**b**) Control experiment in which G-C pair endowed with different dyes are mixed and measured with time. The UV-vis spectra reveals how the acceptor band arises at expense of the donor band until reaching the equilibria 20 hours later. (**c**) Self-sorting experiment in which the UV-vis spectra is the same that the sum of the spectra when these samples are analyzed separately.

4.2 Conclusions.

In this work, we provide what we believe is the first example in which self-sorting self-assembly is ruled *mainly* (for the G-A + A-U pair) or *solely* (for the G-C + iG-iC combination) by chelate cooperativity. A combination of 1D and 2D NMR, CD, and fluorescence spectroscopy with donor-acceptor FRET pairs, in diverse solvents and concentration ranges, clearly confirms that it is the strong propensity of each dinucleoside monomer to form its own cyclic tetramer assembly independently, and not H-bonding complementarity, what drives narcissistic self-sorting. Such supramolecular scenario allowed us to prepare binary or ternary (not shown here) dinucleoside mixtures, sort them in their corresponding H-bonded macrocycles, and address their gradual dissociation independently by increasing temperature or adding a denaturation agent.

4.3. Experimental Section.

The General Methods detailed in the Experimental Section **1.4** of *Chapter 1* and Experimental Section **2.3** of *Chapter 2* are also applicable here.

4.3.1. Synthesis and Characterization.

The synthesis and characterization of **G**, **C**, **A**, **U**, **iG** and **iC** mononucleosides and **G1-B1-C1**, **iG1-B1-iC1**, **A1-B1-C1**, dinucleosides have been reported in *Chapter 2* and *Chapter 3* respectively and in our published works: *Org. Biomol. Chem.* **2015**, *13*, 4506–4513; *Angew. Chem. Int. Ed.* **2015**, *54*, 6780–6784 and *Angew. Chem. Int. Ed.* **2016**, *55*, 223–227. Ref: 158

The synthesis of dyes molecules will be published in the following article, which is under preparation:

"Self-sorting Phenomena Governed by Chelate Cooperativity", C. Montoro-García, D. Serrano-Molina, M. J. Mayoral, D. González-Rodríguez.

4.3.2 NMR and Optical Spectroscopy Dilution and Titration Experiments.

The NMR and optical spectroscopy experiments detailed in the Experimental Section **1.4** of *Chapter 1* and Experimental Section **2.3** of *Chapter 2* and Experimental Section **3.3.2** of *Chapter 3* are also applicable here.

Chapter 5

Impact of Ring Size on Chelate Cooperativity in Noncovalent Macrocyclizations

5.1 Monomer Design

Once the macrocyclization process and self-sorting phenomena of **G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-U1** were studied, *Chapter 5* is devoted to assess which is the impact on the chelate cooperativity when the length of the central block is lengthened or shortened. These modifications of the central block are constituted from a simple C-C bond to finish with a five ring member π-conjugated oligophenylene-ethynylene structure between nucleobases. All of these central blocks are substituted at both edges by the same G-C nucleobase pair (Figure 5.1), because as we explained previously, this base-pairing motif presents an unsymmetric H-bond pattern and thus it enhances the chelate effect and is the best way to compare the stability that each central block confer to the tetrameric assembly. **A_{alk10}-U1** is the only one exception in this trend in order to study the differences between base-pairs.

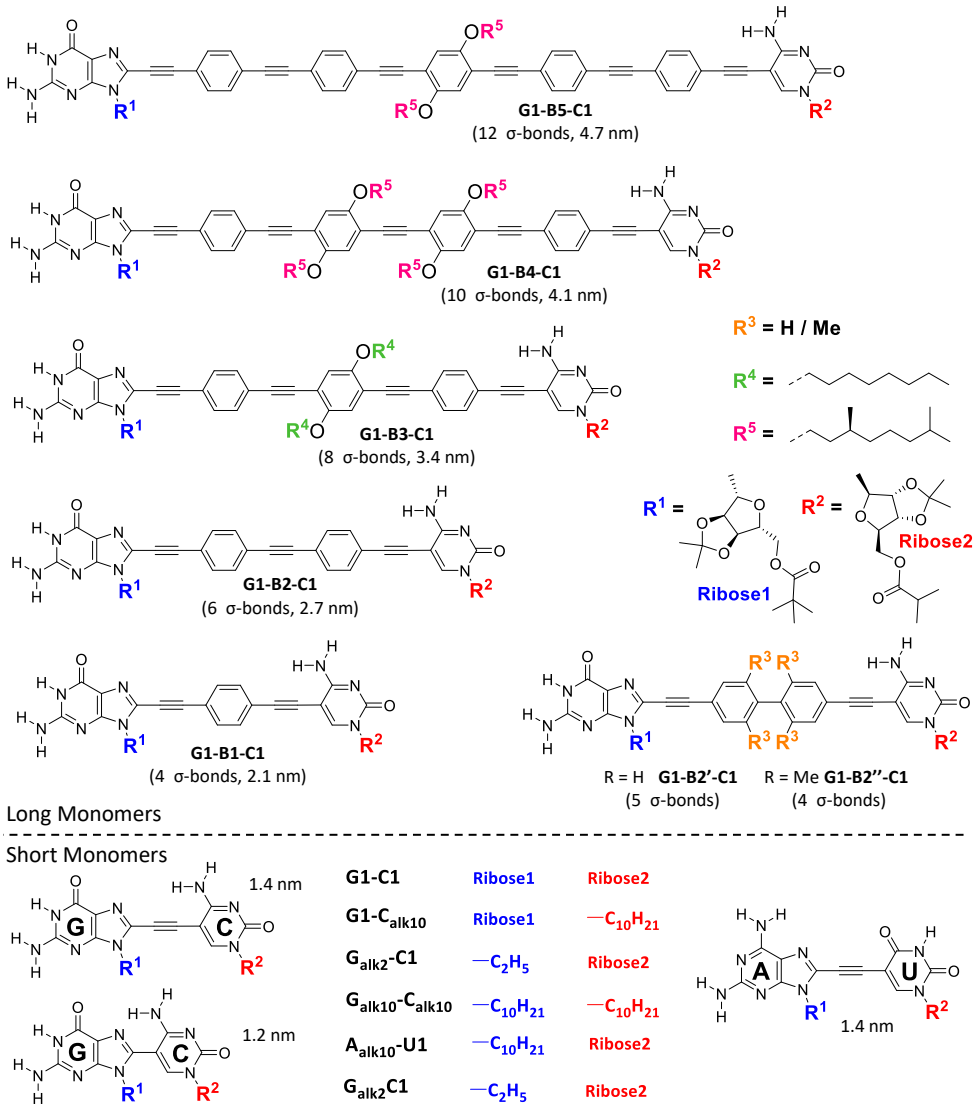


Figure 5.1. Chemical structure of the dinucleoside monomers employed in this work. The number of rotatable σ-bonds in the linking π-conjugated blocks and their length are indicated between brackets.

Once the respective monomers were synthesized and studied, we realized on the huge different behavior between those that carry the shorter ethynyl central block, from those which carry the π-conjugated oligophenylene-ethynylene central block. For this reason, we will explain separately both families of monomers due to the measurements and our results, suggest that the macrocyclization process is affected in different ways. Firstly we will comment the stability of the monomers which are formed by large central blocks since they can be easily comparable with **G1-B1-C1** and follow the same tetramerization process. Secondly, the monomers bearing the shortest central blocks will be studied later, due to its differences with the first family.

5.2 Long Monomers.

5.2.1 Monomer Synthesis.

Seven different linear π-conjugated spacers were evaluated in this first part of the *Chapter 5* (**G1-B1-C1** – **G1-B5-C1**, **G1-B2'-C1** and **G1-B2''-C1**; Figure 5.1, long monomers). The main difference among them is their length and the number and nature of σ-bonds (from 4 to 12) present in their π-conjugated skeleton. The strategy to synthesize all of these unsymmetric G-C monomers has been explained in *Chapter 3*, and involves a first Sonogashira coupling between the dihalogenated central block and 5-ethynylcytidine **C1**, and then a second coupling with 8-ethynylguanosine **G1**.[158]

5.2.2 Self-Assembly into Cyclic Tetramers.

With the exception of the monomers having a biphenyl central block (**G1-B2'-C1** and **G1-B2''-C1**), whose self-assembly characteristics will be described later, all **G1-B2-C1** – **G1-B5-C1** molecules display similar self-association features in chlorinated solvents, as described in *Chapter 3* with **G1-B1-C1**.[158]

^{1}H NMR experiments in $CDCl_3$ (Figure 5.2) revealed in all cases a single set of sharp proton signals corresponding to a species associated *via* G-C Watson-Crick H-bonding interactions. The G amide (H^1) and the C amine (H^2) signals, which are characteristic probes for such H-bonding association, are respectively found at 13.4 and 10.0 ppm (see the projected ^{1}H NMR in Figure 5.2) and display NOE cross-peaks between them. The shape and position of these H-bonded signals in **G1-B1-C1** – **G1-B5-C1** are not very sensitive to temperature or concentration changes within the $5x10^{-2}$ M to 10^{-4} M range studied. This is in sharp contrast to the behavior of the **G**+**C** 1:1 complex, whose G-H^1 and C-H^2 signals slightly broaden and shift upfield with dilution, and suggests that a particularly stable supramolecular entity is formed by the **GC** dinucleosides.

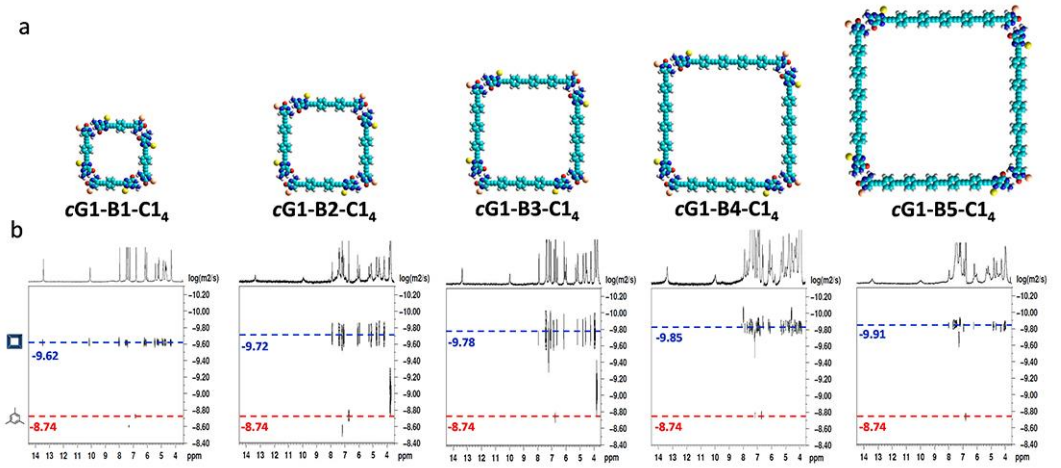


Figure 5.2. (**a**) Cyclic tetramer models and calculated hydrodynamic radii. (**b**) DOSY NMR spectra ($CDCl_3$, 10^{-2} M, 298 K) of **G1-B1-C1** – **G1-B5-C1**. The average diffusion coefficients measured are indicated in blue for **G1-B1-C1** – **G1-B5-C1** and in red for mesitylene, which was employed as an internal reference.

Diffusion NMR experiments were performed with 10^{-2} M **G1-B1-C1** – **G1-B5-C1** solutions in $CDCl_3$ in the presence of equimolar amounts of mesitylene, which served as an internal reference. The DOSY NMR spectra acquired at 298 K, shown in Figure 5.2, revealed for **G1-B1-C1** – **G1-B5-C1** a single species with a narrow dispersity, characteristic of discrete assemblies, and with an average diffusion coefficient that gradually decreased from $10^{-9.62}$ to $10^{-9.91}$ m^2·s, as the monomer length increased from **G1-B1-C1** to **G1-B5-C1**. In short, and as will be further demonstrated below, the experiments performed in this apolar solvent suggest that **G1-B1-C1** – **G1-B5-C1** associate in discrete, H-bonded cyclic tetramer species (*c***G1-B1-C1**$_4$–*c***G1-B5-C1**$_4$).

5.2.3 Cyclic Tetramer Dissociation.

Different methods, in which solvent polarity, concentration / temperature range and monitoring technique are varied, were employed to evaluate qualitatively or quantitatively the thermodynamic stability of the different cyclic assemblies formed by the G-C dinucleosides described in this Chapter. The association behavior of these molecules was again contrasted to the one of a 1:1 mixture of reference mononucleosides **G** and **C**, in which a single Watson-Crick H-bonding interaction is established. Due to the huge difference in thermodynamic stability noted, not every method was suitable to compare all macrocycle sizes.

Variable concentration experiments in THF and DMF.

As stated above, a change in concentration (or temperature) in $CDCl_3$ did not induce appreciable dissociation of the *c***G1-B1-C1**$_4$–*c***G1-B5-C1**$_4$ species by NMR. However, when the solvent polarity was increased to THF and then to DMF, G-C H-bonding disruption by competition with the solvent became evident.

As shown in Figure 5.3, diluting the 1:1 **G**+**C** complex in THF-d_8 produced an upfield shift of the G-H^1 and C-H^2 H-bonded proton signals as the population of the Watson-Crick pair decreases at the expense of the solvated **G** and **C** mononucleosides. When the same dilution

experiment was performed with dinucleosides **G1-B1-C1** – **G1-B3-C1**, however, the shape and position of those signals did not experience any significant change. This does not mean that the assemblies are not being dissociated in these experiments. As a matter of fact, the G-H^1 monomer signal was found to be very broad broad in THF-d_8, so the release of increasing amounts of monomer as the concentration is decreased is better monitored in the ribose proton region. As opposed to the 1:1 **G**+**C** complex, **G1-B1-C1** – **G1-B3-C1** monomers are in slow exchange in the NMR timescale with its respective cyclic tetramers, which highlights the singular kinetic stability of this assembly and helps in calculating their molar fraction. Interestingly, the largest macrocycles, **G1-B4-C1** and **G1-B5-C1**, manifested a markedly different behavior. As shown in Figure 5.3, **G1-B4-C1** seems to recover the fast exchange characteristics of the **G**+**C** mixture and display G-H^1 and C-H^2 H-bonded proton signals that shift upfield upon dilution. On the other hand, **G1-B5-C1** exhibits an intermediate behavior and, although the spectra in THF did not present a good resolution, revealed at least two main types of G-H^1 signals. The first one, attributed to the kinetically stabilized **G1-B5-C1** cyclic tetramer, does not change in position (13.4 ppm) with dilution, while the second one shifts downfield in the 13-10 ppm range, which is characteristic of a rapidly exchanging mixture of monomer and H-bonded oligomers. Please note the resemblance of this response to concentration changes with the one shown by **A1-B1-U1** in $CDCl_3$, shown and discussed in *Chapter 3*.[158]

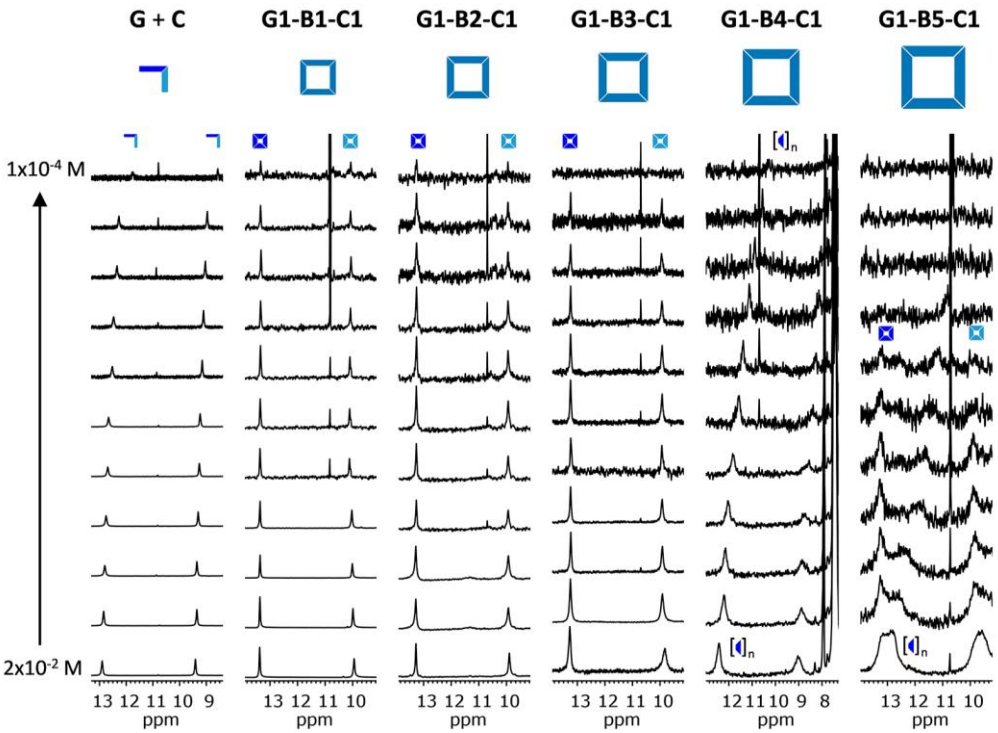


Figure 5.3. *Concentration-dependent experiments in THF-d_8.* Downfield region of the ^{1}H NMR spectra of a 1:1 **G**+**C** mixture and of **G1-B1-C1** – **G1-B5-C1** dinucleosides at different concentrations at 298 K in THF-d_8 showing the G-H^1 (dark blue) and C-H^2 (light blue) proton signals (the signal at 10.7 ppm corresponds to the BHT stabilizer).

These concentration-dependent measurements were used to determine the values of *EM* and K_T in THF-d_8 as will be commented later.

When moving to the more polar DMF-d_7, only *c***G1-B1-C1**$_4$ and, to a lower extent *c***G1-B2-C1**$_4$, showed signs of resisting this aggressive environment for H-bonded species (see Figure 5.4 and also the corresponding temperature-dependent experiments in Figures 5.5, 5.6, and 5.7). It is notable to see again that the cyclic tetramers are in the slow exchange regime with their respective monomers in this polar solvent and exhibit strong all-or-nothing characteristics. This is evidenced by the fact that the shape and position of the G-H^1 tetramer and monomer signals do not experience significant changes with concentration (neither with temperature; Figure 5.7). All the other longer **G1-B3-C1** – **G1-B5-C1** compounds exist as monomeric species in DMF, even at the highest concentrations or the lowest temperatures tested, which denotes an inferior stability of the cyclic systems.

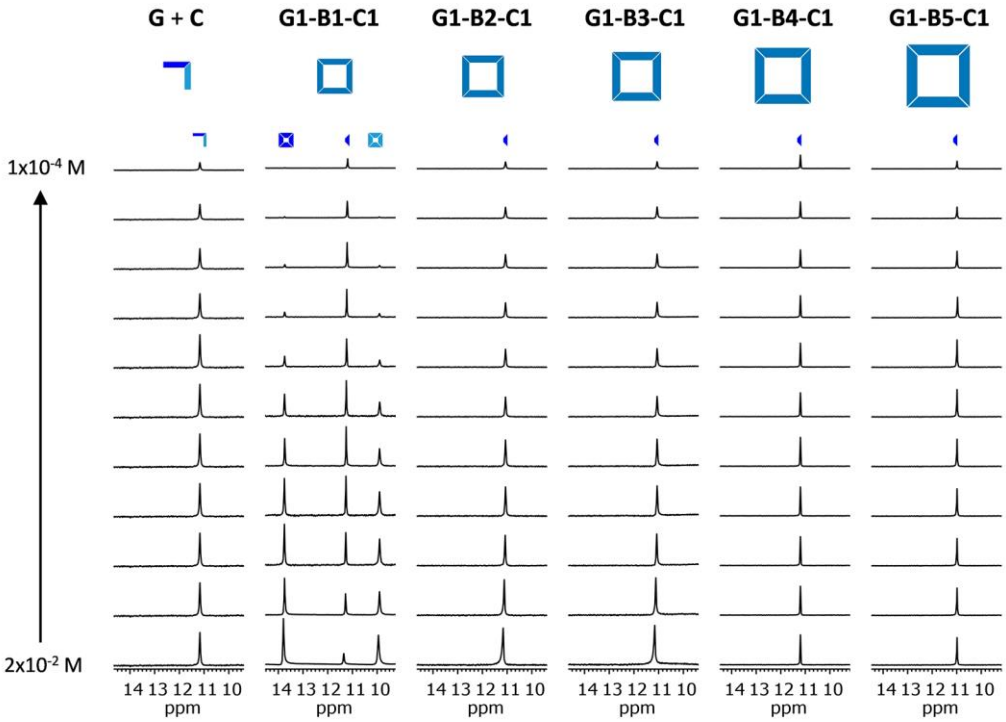


Figure 5.4. *Concentration-dependent experiments in DMF-d_6.* Downfield region of the ^{1}H NMR spectra of a 1:1 **G+C** mixture and of **G1-B1-C1** – **G1-B5-C1** dinucleosides at different concentrations at 298 K in DMF-d_7, showing the G-H^1 (dark blue) and C-H^2 (light blue) proton signals.

Variable temperature experiments in $CDCl_2CDCl_2$, THF and DMF.

Several ^{1}H NMR spectra were also acquired for **G1-B1-C1** – **G1-B5-C1** solutions within the 298-403 K ($CDCl_2CDCl_2$; 10^{-2} M), 273-333 K (THF-d_8; $5x10^{-4}$ M), or 213-328 K (DMF-d_7; 10^{-2} M) temperature ranges. These three solvents constitute a set of environments with very different polarity and ability to compete for H-bonding between nucleosides. As a general rule, association equilibria are shifted to the cyclic tetramer in the less polar $CDCl_2CDCl_2$, and to the monomer in the strongly polar DMF solvent.

At very high temperatures in apolar solvents like $CDCl_2CDCl_2$, the G-C H-bonded signals undergo a broadening and then a gradual upfield shift (Figures 5.5) as the degree of H-bonding is diminished. It can be clearly observed that the longer the monomer, the lower the temperature at which such broadening and upfield shift is detected.

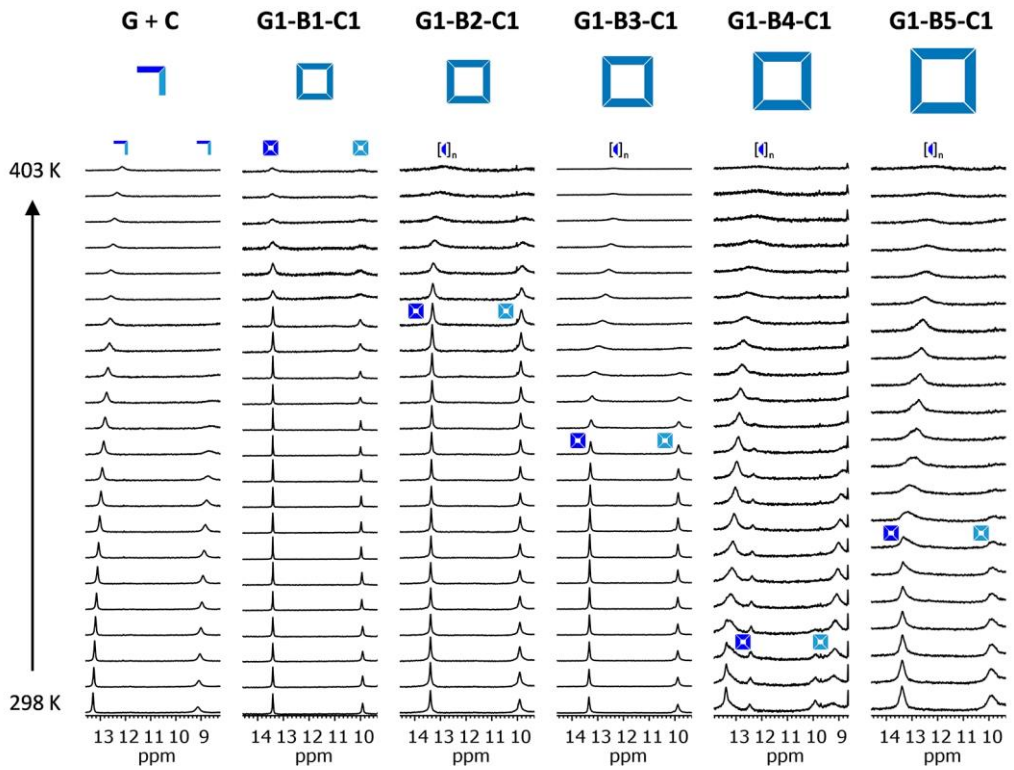


Figure 5.5. *Temperature-dependent experiments in $CDCl_2CDCl_2$.* Downfield region of the ^{1}H NMR spectra of a 1:1 **G**+**C** mixture and of **G1-B1-C1** – **G1-B5-C1** dinucleosides at different temperatures in $CDCl_2CDCl_2$ (*C* = $1.0x10^{-2}$ M) showing the G-H^1 (dark blue) and C-H^2 (light blue) proton signals.

More polar solvents like THF or DMF offer a different scenario. As the temperature is increased, the slowly exchanging *c*GC_4 proton signals gradually disappear at the expense of the GC monomer signals (see Figures 5.6 and 5.7). However, the temperature at which the *c*GC_4 species is dissociated was found to drastically depend again on 1) solvent polarity and 2) monomer length. Such temperature-dependent association changes in THF-d_8 and DMF-d_7 are respectively plotted in Figure 5.13. Cyclic tetramer stability is clearly higher for the shorter monomers, which are formed quantitatively in THF, and lower for the longer **G1-B3-C1** – **G1-B5-C1** monomers, whose self-association, as explained above, is fully hampered in the polar DMF solvent and only the monomer is detected in the whole temperature range. These temperature-dependent measurements were used to determine the values of Δ*S* and Δ*H* in THF-d_8 and DMF-d_6 as will be commented later.

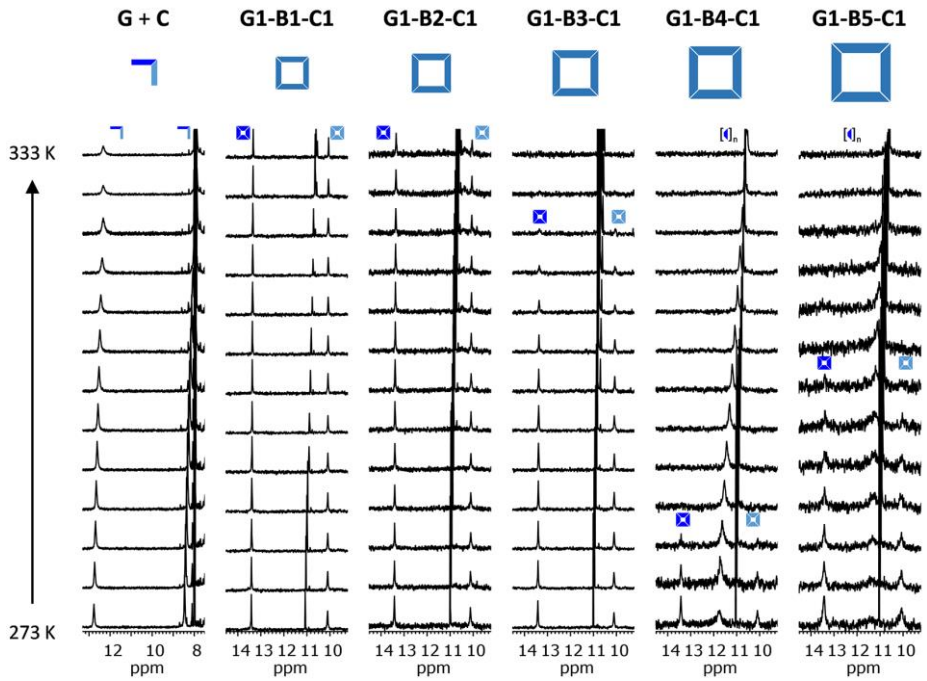


Figure 5.6. *Temperature-dependent experiments in THF-d_8.* Downfield region of the ^{1}H NMR spectra of a 1:1 **G**+**C** mixture and of **G1-B1-C1** – **G1-B5-C1** dinucleosides at different temperatures in THF-d_8 (C = 1.0x10^{-2} M) showing the G-H^1 (dark blue) and C-H^2 (light blue) proton signals (the signal at 10.7 ppm corresponds to the BHT stabilizer).

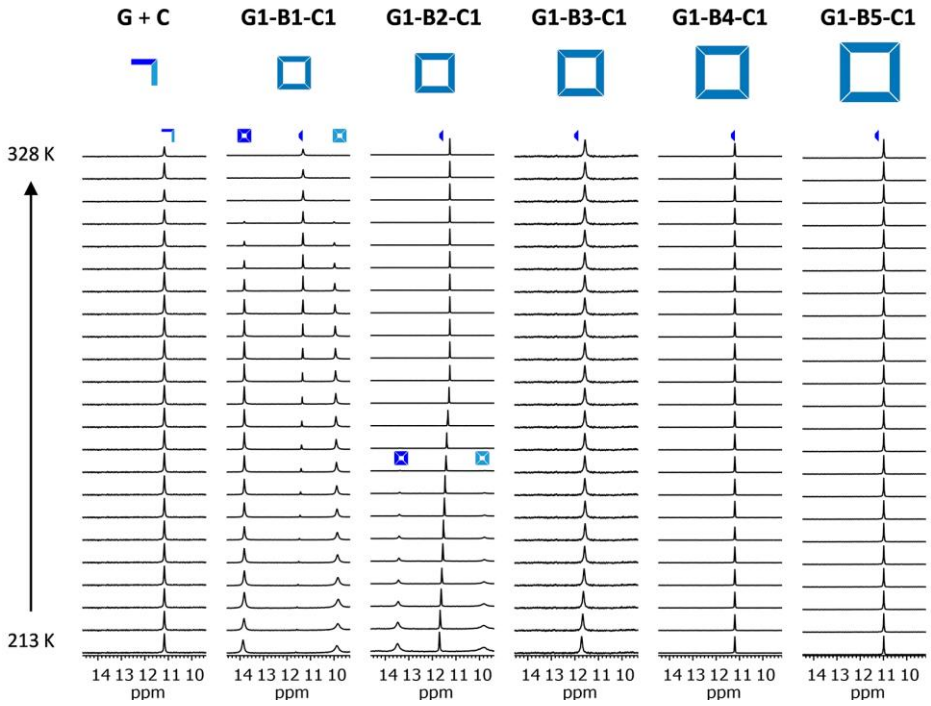


Figure 5.7. *Temperature-dependent experiments in DMF-d_7.*Downfield region of the ^{1}H NMR spectra of a 1:1 **G**+**C** mixture and of **G1-B1-C1** – **G1-B5-C1** dinucleosides at different temperatures in DMF-d_7 (C = 1.0x10^{-2} M) showing the G-H^1 (dark blue) and C-H^2 (light blue) proton signals.

The experiments in THF at low concentrations (*ca.* 10^{-4} M) were contrasted with more sensitive CD and absorption spectroscopy techniques. We have previously determined that dinucleoside association can be spectroscopically monitored by an absorption and emission red-shift and by the appearance of characteristic CD signals. The evolution of these optical features with sample concentration and temperature nicely matches the cyclotetramerization process observed by NMR. CD and UV-vis spectra were recorded for **G1-B1-C1** – **G1-B5-C1** in THF at a $5x10^{-4}$ M concentration within the 273-328 K temperature range. The results are collected in Figures 5.8a and 5.8b.

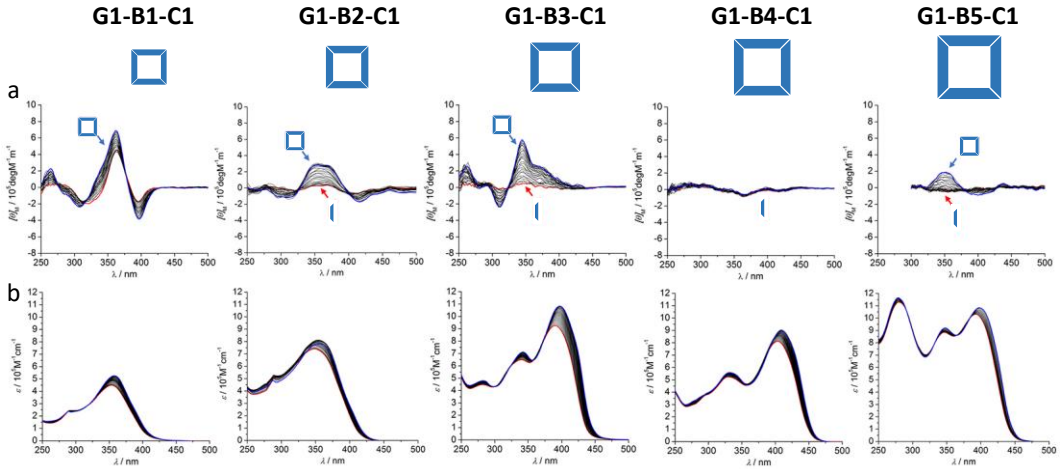


Figure 5.8. Temperature-dependent (**a**) CD and (**b**) UV-vis experiments of **G1-B1-C1** – **G1-B5-C1** dinucleosides in THF at $1.25 \cdot 10^{-4}$ M (273-328 K).

These data complement the temperature-dependent NMR data and support the trends observed at higher concentrations in THF: *the longer the dinucleoside monomer the weaker the persistence of the corresponding cyclic species at high temperatures*.

Denaturation experiments. Dissociation in $CDCl_3$ with DMSO-d_6 and with C mononucleoside.

Probably the most revealing set of experiments were those in which the cyclic assemblies were denatured by gradually adding a substance that competes for H-bonding. Concretely, cyclic tetramer dissociation was monitored by ^{1}H NMR in 10^{-2} M $CDCl_3$ solutions of **G1-B1-C1** – **G1-B5-C1** by adding increasing amounts of either DMSO-d_6 or the complementary pyrimidine **C** nucleoside (Figures 5.9 and 5.11).

Figure 5.9 displays the evolution of the G-H^1 and C-H^2 region of the NMR spectra as the volume fraction of DMSO-d_6 is increased in $CDCl_3$-DMSO-d_6 mixtures at constant **G1-B1-C1** – **G1-B5-C1** dinucleoside concentration. These experiments, on one hand, further supported quantitative cyclic tetramer formation in apolar solvents like $CDCl_3$ and, on the other, provided a qualitative measure to compare cycle stability, which is progressively dissociated in the presence of this polar H-bonding-competing cosolvent.

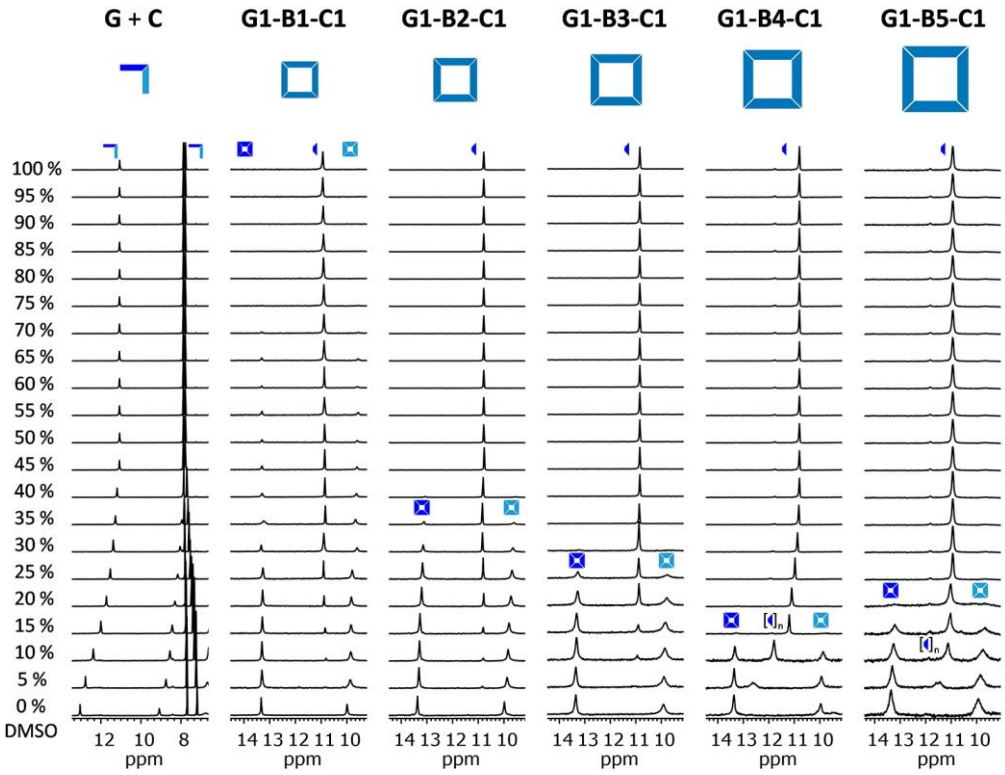


Figure 5.9. *Denaturation experiments with increasing amounts of DMSO.* Downfield region of the ^{1}H NMR spectra, showing the H-bonded G-amide (H^1; dark blue) and C-amine (H^2; light blue) proton signals, of a 1:1 **G+C** mixture and of **G1-B1-C1** – **G1-B5-C1** dinucleosides as the volume fraction of DMSO-d_6 is increased in $CDCl_3$-DMSO-d_6 mixtures at C = 1.0x10^{-2} M and T = 298 K. Additional spectra were recorded at low DMSO content for **G1-B2-C1** – **G1-B5-C1**, but are not shown here for the sake of clarity and homogeneity.

Similar to the behavior noted in DMF-d_7, as the macrocycle is dissociated with DMSO-d_6 a new signal at 10.9 ppm grows at the expense of the original *c*GC_4 signals, which is attributed to the solvent-bound G-H^1 amide proton in the monomeric GC species. It is interesting to remark, as we concluded in *Chapter 3*,[158] that the cyclic tetramer constitutes in all cases a thermodynamically and kinetically stabilized species in the overall self-assembly landscape. Hence, in these and the rest of NMR experiments in which this cyclic assembly is dissociated in polar solvents, we typically detect two sets of signals in slow exchange that are assigned to *c*GC_4 and to the GC monomer. This behavior, observed for all **G1-B1-C1** – **G1-B5-C1** dinucleosides, sharply differs from the one expected for a supramolecular polymerization process (see below) or from the one observed for the 1:1 mixture of **G** and **C** mononucleosides (also represented in Figure 5.9), where H-bonded and monomer species are in fast exchange and, as a result, a gradual upfield shift of the H-bonded signals is monitored.

The molar fraction of *c*GC_4, obtained by integration of several isolated C-H proton signals, is represented in Figure 5.10a as a function of the DMSO-d_6 volume fraction for the different **G1-B1-C1** – **G1-B5-C1** systems studied. A clear general relationship between cyclic tetramer size and DMSO content was observed that is in line with the stability trends determined from the previous concentration- and temperature-dependent experiments: *the smaller the macrocycle, the higher its resistance to this strongly competing cosolvent*.

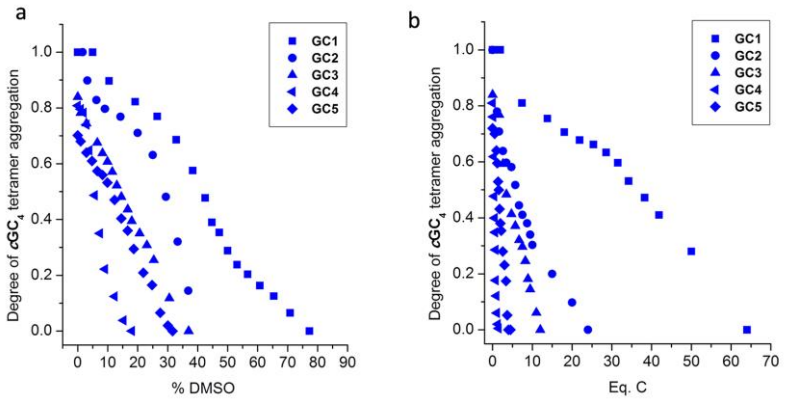


Figure 5.10. Evolution of the degree of *c***GC**$_4$ association as a function of (**a**) DMSO-d_6 volume fraction and (**b**) equivalents of **C**.

But furthermore, a distinct cooperative all-or-none behavior can be noted by inspecting the set of spectra in Figure 5.9. One can clearly see that the position of the G-H^1 proton signal H-bonded to C in the cyclic tetramer, at *ca.* 13.4 ppm, does not change with the DMSO content. On the other hand, the monomer G-H^1 signal, at *ca.* 10.9 ppm, does not change in shape or position for **G1-B1-C1**, **G1-B2-C1** or **G1-B3-C1**, but it does change for **G1-B4-C1** and **G1-B5-C1**. Concretely, in these last cases this signal shifts upfield and narrows with increasing DMSO content, which indicates that the **G1-B4-C1**/**G1-B5-C1** monomers are actually in fast equilibrium with non-cyclic oligomeric species and, therefore, that the strong all-or-none cooperative features exhibited by the shorter dinucleosides are gradually lost when increasing monomer length.

The different stability of the **G1-B1-C1** – **G1-B5-C1** cyclic tetramers was also evaluated through denaturation experiments in which increasing amounts of a **C** nucleoside stopper were gradually added to **G1-B1-C1** – **G1-B5-C1** solutions in $CDCl_3$ (see Figure 5.11). In these experiments, the mononucleoside competes with the dinucleoside for binding to the complementary base, and the titration with **C** progressively transforms the *c*GC$_4$ tetramolecular macrocycle into the C·GC bimolecular complex (please come back to Section 3.1.4 in *Chapter 3* to further details).[158] Since *c*GC$_4$ is a kinetically stabilized species in the NMR timescale, both *c*GC$_4$ and C·GC afford separate signals in slow exchange, whereas the excess of **C** mononucleoside in solution is in fast exchange with C·GC. As can be observed in Figure 5.11, as increasing amounts of **C** are added, a new G-H^1 proton signal, corresponding to the H-bonded C·GC complex, arises at slightly downfield chemical shifts (*i.e.* at around 13.5 ppm) and increases in intensity at the expense of the *c*GC$_4$ signals.

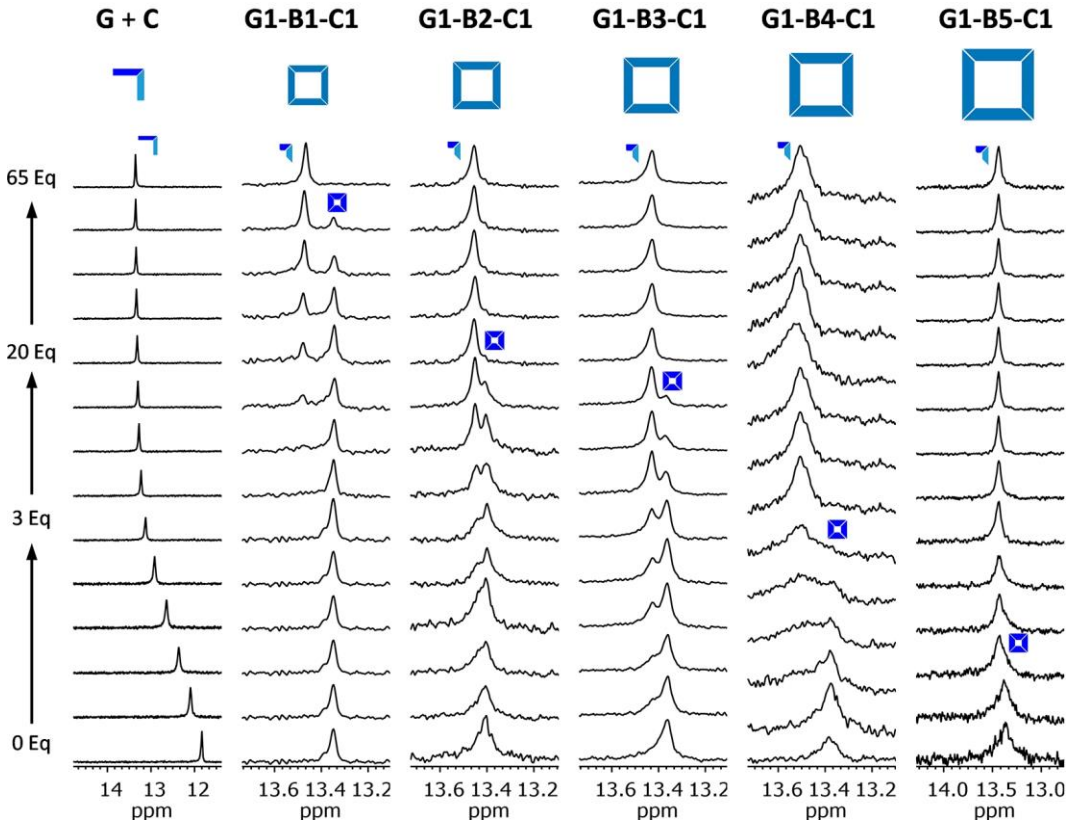


Figure 5.11. *Denaturation experiments with increasing amounts of C mononucleoside.* Changes observed in the G-H^1 region of the ^{1}H NMR spectra of **G** and of **G1-B1-C1 – G1-B5-C1** (C = 1.0x10^{-2} M and T = 298 K) by adding increasing amounts of **C**. Additional spectra were recorded at low equivalents of **C** added for **G1-B2-C1 – G1-B5-C1**, but are not shown here for the sake of clarity and homogeneity.

The G-C binding interaction is the same in both *c*GC$_4$ and C·GC assemblies, as well as for all *c***G1-B1-C1**$_4$–*c***G1-B5-C1**$_4$ ring sizes. However, the amount of competing mononucleoside that is needed to observe the total disappearance of the *c*GC$_4$ ^{1}H NMR signals is very different and considerably decreases as the monomer length increases. Specifically, the molar fraction of *c*GC$_4$ for all **G1-B1-C1** – **G1-B5-C1** monomers is represented in Figure 5.10b as a function of the equivalents of **C** added. In these experiments we are actually monitoring the competition between intra- and intermolecular binding, and therefore these plots and the underlying quantitative analysis provide a direct evaluation of the *EM* of a cyclic system as explained in CDCl$_3$.[158]

Thermodynamic and Kinetic Parameters.

The results obtained in these competition tests not only support the clear conclusion obtained from all the experiments performed to assess and compare the stability of the cyclic tetramers formed by **G1-B1-C1** – **G1-B5-C1**: *the longer the rigid π-conjugated block connecting the complementary bases, the lower the thermodynamic stability of the cyclic assembly*. They also afford an indication that this stability trend stems from the *EM* of each particular system and, thereof, from the different strength of the chelate effect. Table 5.1 compiles the *EM*

values calculated for **G1-B1-C1** – **G1-B5-C1** in THF and $CHCl_3$ in the different dilution or competition experiments performed (see *Chapter 3* for further details).[158]

Solvent K_{ref} / M[1]	**M**	***EM***[a] M	**Δ*H*** kJmol^{-1}	**Δ*S*** Jmol^{-1}K^{-1}
DMF	**G1-B1-C1**	2.2x10^{2} [b]	-155.2	-425.0
5.7±0.3	**G1-B2-C1**		-166.3	-558.8
THF	**G1-B1-C1**	2.0x10^{2} [b]	-98,7	-32,3
1.5±0.1x10^{3}	**G1-B2-C1**	2.4x10^{0}	-91.9	-66,3
	G1-B3-C1	1.6x10^{-1}	-95.8	-87,6
	G1-B4-C1			
	G1-B5-C1	1.2x10^{-3}	-101,6	-159,8
$CHCl_3$	**G1-B1-C1**	9.1x10^{2} [b]		
2.8±0.3x10^{4}	**G1-B2-C1**	1.1x10^{1}		
	G1-B3-C1	4.9x10^{-1}		
	G1-B4-C1	3.1x10^{-2}		
	G1-B5-C1	2.2x10^{-3-3}		

Table 5.1. Cyclotetramerization constants (K_T), reference intermolecular association constants (K_{ref}) and effective molarities (*EM*) obtained for **GC**/**iGiC**/**AU** from different experiments. [a] Determined as: $EM = K_T/K_{ref}^4$ using the data calculated from the NMR dilution in DMF-d_6 (Figure 5.4), NMR dilution in THF-d_8 (Figure 5.3), NMR competition experiments with **C** in $CDCl_3$ (Figure 5.11). [b] *EM* values ranging between 2.2x10^2-3.6x10^2 M (DMF), 1.8x10^2-7.3x10^2 M (THF), or 8.1x10^2-9.1x10^2 M ($CHCl_3$) had been previously determined by us for **G1-B1-C1** using other methods.[156]

The calculated K_{ref} and *EM* values were then used to simulate speciation profiles for each dinucleoside molecule in DMF, THF and $CHCl_3$ (Figure 5.12) These curves relate the concentration of each supramolecular species with total concentration and are able to reproduce quite satisfactorily the dissociation behavior observed for *c***G1-B1-C1**$_4$–*c***G1-B5-C1**$_4$ in the dilution experiments in THF (Figure 5.12b), where both simulated and experimental results are combined.

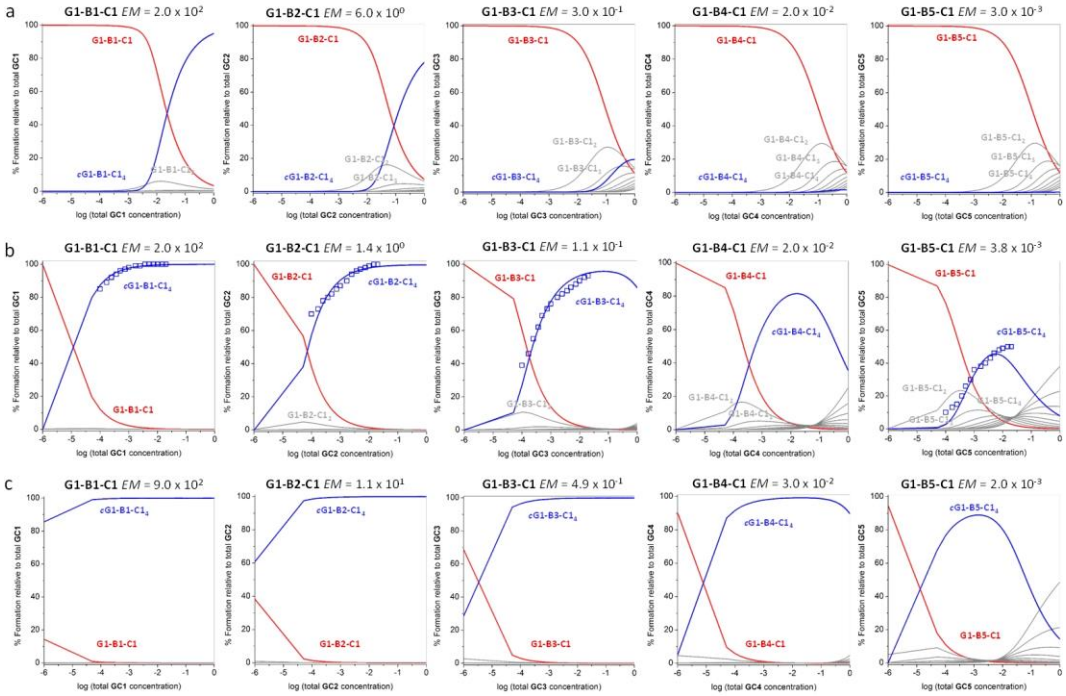


Figure 5.12. Simulated speciation curves (lines) and experimental dilution data indicating the degree of c$\mathbf{M}_4$ association of **G1-B1-C1** – **G1-B5-C1** in (**a**) DMF, (**b**) THF, and (**c**) $CHCl_3$.

As the macrocycle becomes larger, the magnitude of *EM* experiences in both solvents a drastic decrease that encompasses 5 orders of magnitude, from over 10^2 M for **G1-B1-C1** to 10^{-3} M for **G1-B5-C1**. Since the binding interaction that sustains the cyclic assemblies is the same in all cases: G-C Watson-Crick pairing, a weaker chelate cooperativity is identified here as the main cause for the notable reduction in thermodynamic stability observed for the larger cycles. To further support this hypothesis, the temperature-dependent NMR and CD experiments in THF were analyzed to determine the enthalpic (ΔH) and entropic (ΔS) changes of the cyclotetramerization process.[158] The corresponding van't Hoff plots are shown in Figure 5.13. Parallel lines were obtained that manifest that the enthalpy of this cyclization process is very similar for all **G1-B1-C1** – **G1-B5-C1** monomers and that entropy is the actual responsible for the stability differences observed.

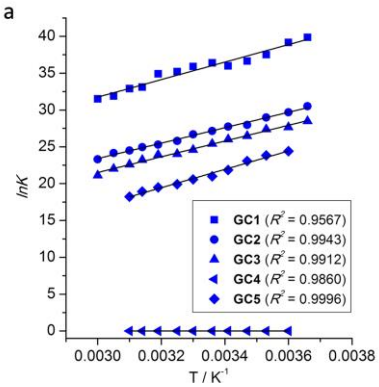


Figure 5.13. Van't Hoof plot of **G1-B1-C1** – **G1-B5-C1** dinucleosides at different temperatures in THF measured in (**a**) NMR (C = 1.0x10^{-2} M) and (**b**) CD (C = 1.25x10^{-4} M).

In order to rationalize the entropic origin in the reduction of *EM*, let us focus on the cyclization event and compare open and cyclic tetramer species, as it is schematically shown in Figure 5.14.

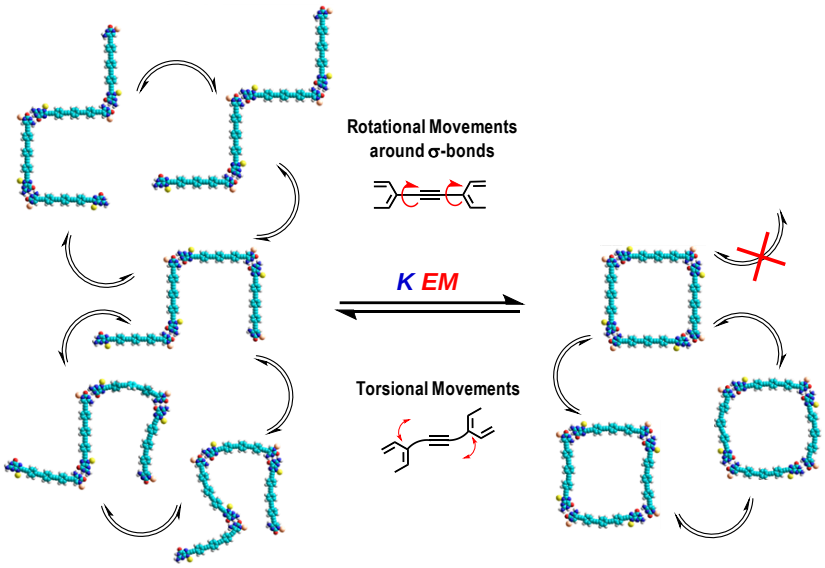


Figure 5.14. Suggested cyclization events when the monomer becomes larger (**G1-B1-C1**).

As stated in the introduction, the gain in stabilization when going from an open to a cyclic system (the magnitude of the chelate effect), is represented by the product *K·EM*, where *K* is the reference G-C association constant, since there is an additional binding event to form the cycle and it is the same for all macrocycles independently of their size. *EM* is the factor that takes into account that this last binding event to form the cycle is intramolecular and different from the rest. Usually, *EM*s are high, and thus cycle formation is favored, when there is no significant strain generated upon cyclization. In our case, all **G1-B1-C1** – **G1-B5-C1** monomers share a rigid structure that was designed to produce cyclic squared-shaped assemblies devoid of strain. This is demonstrated by the fact that cyclization is not associated with large enthalpic differences between the different monomers. However, we should take

into account other issues that affect the entropic term in *EM* and that are related with the *degrees of freedom that are lost upon cyclization*.

Let's first consider the rotation around σ-bonds in the oligo(ethynylene-phenylene) block connecting the G and C bases. These rotations are not restricted upon cyclization, all σ-bonds can still rotate freely in the cyclic species. However, there are some conformations related to the relative disposition of the nucleobases that cannot be accessed in the cycles. As shown in Figure 5.15, rotation around these bonds in the open oligomers can produce multiple conformations in which the Watson-Crick edges alternate between "*syn"* and "*anti"* relative arrangements. However, cyclization demands the Watson-Crick edges to arrange exclusively in a "*syn"* relative conformation. This is a degree of freedom that is lost when comparing cyclic and open species, and must contribute to an entropic reduction in the maximum attainable *EM* of the cyclic system.

Let's now consider torsional movements, which can be accessed by stretching and bending of (mainly) the σ-bonds in the relatively flexible ethynylene-phenylene skeleton. Comparing again a linear and a cyclic species it is clear that these collective torsional movements should be considerably restricted in the more rigid cyclic structure, which presents an additional binding site, than in the more flexible linear oligomers, which enjoy free end-groups.

In short, when going from an open to a closed species the number of degrees of freedom associated with rotational and torsional movements of (mainly) σ-bonds is reduced. This would explain why EM values drop significantly as the length of the oligo(ethynylene-phenylene) central block increases, and hence the number of σ-bonds that can rotate and torsion.

The Special Case of G1-B2'-C1 and G1-B2''-C1: Supramolecular Polymerization.

Compounds **G1-B2'-C1** and **G1-B2''-C1** displayed an anomalous behavior in all of the experiments described so far that made us not include them in our general experimental trends and dedicate them a separate section. Biphenyl spacers with a different degree of steric hindrance were installed as central blocks in these compounds. Hence, they were actually synthesized to analyze the effect of partially freezing or totally blocking a rotor in the spacer, while maintaining a similar distance between the complementary nucleosides at the edges.

The results obtained for **G1-B2'-C1** actually fit very well with our overall analysis although they do not correlate in the trends plotted in Figure 5.10. This dinucleoside reveals the same supramolecular features as compounds **G1-B1-C1** – **G1-B5-C1** in the different experiments exposed in this Chapter: it forms robust cyclic tetramers in apolar solvents that dissociate with the addition of polar cosolvents, leading to a monomer-cyclic tetramer equilibrium in slow exchange at the NMR timescale. Quantitatively, in terms of thermodynamic stability and *EM* values, it can be compared with **G1-B2-C1**, despite **G1-B2'-C1** has 5 σ-bonds instead of 6 and in one of them, markedly different from the aryl-ethynyl sigma bond that is common to **G1-B1-C1** – **G1-B5-C1**, rotation is sterically hindered to a small extent. Therefore, this monomer is a direct proof that not only the number, but also the nature of the rotors should be considered when making an analysis of the possible constituents of the central block are chosen.

In contrast, **G1-B2''-C1** is the only one that discloses an anomalous behavior that is consistent with the formation of supramolecular polymers: 1) ^{1}H and DOSY NMR experiments

in $CDCl_3$ revealed very broad signals characteristic of a distribution of H-bonded oligomers; 2) addition of DMSO-d_6 or dilution experiments in polar solvents such as THF-d_8 or DMF-d_7 resulted in a gradual upfield shift of the H-bonded proton signals, very similar to the one monitored for the 1:1 **G**+**C** mixture, which sharply contrasts the all-or-nothing features observed for the other monomers; 3) while **G1-B1-C1** – **G1-B5-C1** yielded clear solutions in any of the solvents studied, **G1-B2''-C1** formed viscous gels in $CHCl_3$, CH_2Cl_2 and chlorobenzene.

Compound **G1-B2''-C1** is equipped with a biphenyl spacer with 4 methyl substituents that impede rotation around the central sigma bond, therefore orienting the phenyl groups in orthogonal planes. Hence, this block should be seen as a completely rigid central block, revealing the same stability that **G1-B1-C1**. In contrast, this is not the behavior showed by this **G1-B2''-C1** monomer revealing several problems to dissolve it apolar solvents and the apparition of the monomeric species in more polar solvents as THF, DMF or DMSO. In this case, we suggest that the conjugation between nucleobase through the triple bond with one of the rings of the central block is quite high. This conjugation guide nucleobases to point the Watson-Crick recognition sites in an orthogonal angle, being thus an additional impediment to reach the flat "*syn*" conformation to lead the process to ring-closure.

5.3 Short Monomers.

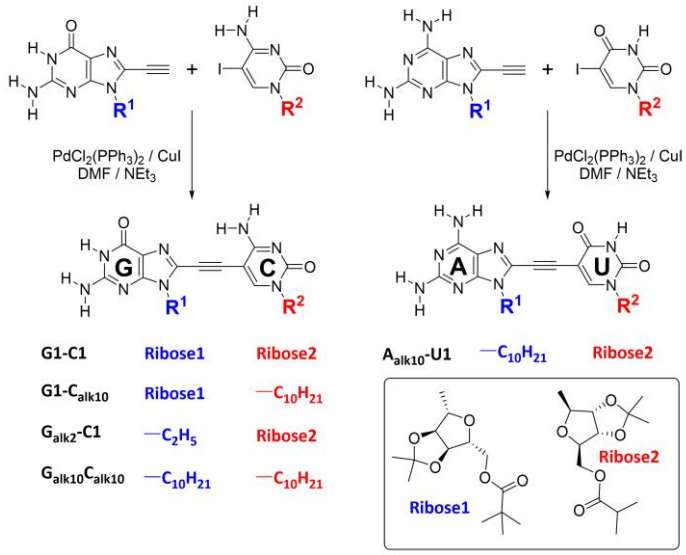
The extrapolation in this trend to small monomers would then induce to think that shorter spacers (see also Scheme 5.1) can lead to an enhancement of chelate cooperativity, and thus to an increased cyclic tetramer stability. However, we reasoned that short distances between nucleobases would also bring new effects that were not dominant or absent in the larger monomers. On one hand, steric effects between bulky lipophilic ribose substituents should begin to manifest as the nucleobases come closer. Steric hindrance is much higher in the "*syn*" conformation, the one required for cyclization, so we should expect that chelate cooperativity would be strongly impacted. On the other hand, as the length of the π-conjugated linker becomes shorter, one would expect that the electronic communication between nucleobase heterocycles will become more important. This might produce other kind of cooperative/anticooperative effects in which the binding of one Watson-Crick edge in the monomer would affect *electronically* the binding strength at the opposite edge.[165]

The impact of these two effects, steric and electronic, on chelate cooperativity is analysed here in shortly spaced ethynylene-linked dinucleoside monomers. For such goal, 5 different monomers were synthesized and studied both experimentally in solution, to analyse their self-assembly characteristics, and with computational DFT methods. One of these monomers contain the 2-aminoadenine (A)-Uracil (U) pair (**A_{alk10}-U1**) and four of them the guanine (G)-cytosine (C) pair (**G1C1**, **G1-C_{alk10}**, **G_{alk2}-C1**, **G_{alk10}-C_{alk10}**), and differ in the substituents placed at the purine N-9 and at the pyrimidine N-1, as shown in Scheme 5.1.

5.3.1 Synthetic Strategy to Unsymmetric Short Monomers.

The synthetic strategy toward unsymmetric monomers have been explained previously in Section 3.1 of *Chapter 3*. However, the synthetic strategy to obtain the unsymmetric monomers linked through an ethynylene spacer is even more straightforward. In this case, keep some amount of the respective nucleobase in the halogenation step is required. Sonogashira reaction between the halogenated (pyrimidine) and ethynylated (purine) derivatives were carried out to afford the desired monomer. Guanosine should contain the ethynyl group due to if the Sonogashira reaction is carried out with the halogenated derivative, the protection of the carbonyl group is necessary (please come back to *Chapter 1* for further explanations) and finally it must be removed in a subsequent reaction with TBAF. This synthetic pathway was not employed due to the value of this compounds and the addition of one more step in the synthetic route.

[165] The electronic effects in the non-bonded nucleobase when a Watson-Crick H-bond is formed in the other side of the molecule is being investigated in collaboration with Prof. Dr. Célia Fonseca Guerra, Amsterdam Center for Multiscale Modeling (ACMM) at the University of Leiden.



Scheme 5.1. General structure and synthesis of ethynylene-bridged G-C monomers and A-U monomer.

5.3.2 Macrocyclization Process of Short Monomers.

With the aim of maintaining the same ribose substitution pattern as in our previous study, **G1-C1** was firstly synthesized (see Figure 5.1 and Scheme 5.1). When the last Sonogashira coupling reaction was carried out in THF, a gel was formed after a few hours. Since the use of DMF in this reaction avoided gel formation we reasoned that THF gelation was caused by the generation of long supramolecular polymers *via* Watson-Crick H-bonding between the terminal G and C in the final dinucleoside product. This behaviour strongly contrasts with the one shown by the longer monomers (Figure 5.1) of identical ribose substitution pattern studied previously, which always afforded clear non-viscous solutions. Therefore, after product purification and characterization, **G1-C1** was subjected to a number of additional studies.

We first investigated if the ability of **G1-C1** to gelate non-polar solvents was a general property. We found out that **G1-C1** is either insoluble (CH_3OH, CH_3CN, acetone, diethyl ether, cyclohexane, and heptane) or forms strong gels (toluene, benzonitrile, THF, $CHCl_3$, $CHCl_2CHCl_2$ and chlorobenzene) after heating and cooling back to room temperature in a wide variety of organic solvents, excluding the very polar DMF or DMSO. The microstructure of the gels formed in chlorobenzene and benzonitrile was further examined by scanning electron microscopy (SEM; Figure 5.16). The SEM images revealed fibers that are connected and entangled over all the surface, which is characteristic of organogels.

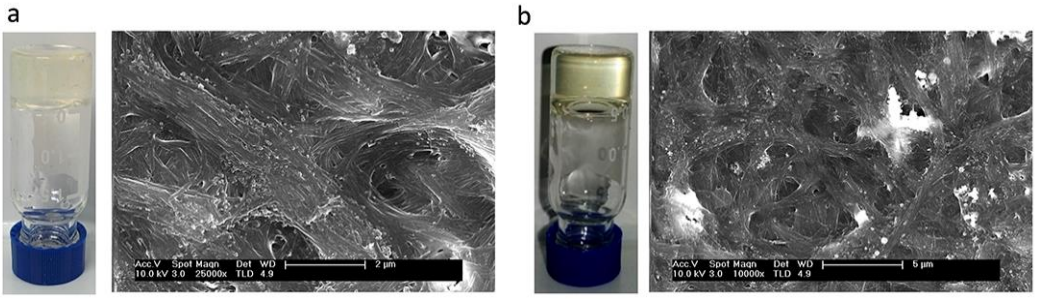


Figure 5.15. Inverted vial pictures and SEM images of **G1-C1** in (**a**) chlorobenzene and (**b**) benzonitrile.

Therefore, the supramolecular behaviour and self-assembled morphology of **G1-C1** is different from the longer monomers. A first inspection of the computational molecular models (Figure 5.16) clearly reveals that steric hindrance between riboses at **R1** and **R2** in the "*syn*" conformation is high enough to almost exclusively populate the "*anti*" conformation. This conformational restriction impedes the formation of cyclic species and the bases instead interact producing long-H-bonded polymers.

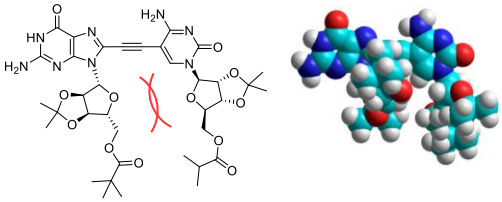


Figure 5.16. Hyperchem model showing the steric hindrance between ribose moieties of **G1-C1** monomer.

We then decided to synthesize 3 additional G-C monomers (**G1-C$_{alk10}$**, **G$_{alk2}$-C1** and **G$_{alk10}$-C$_{alk10}$**) in which steric hindrance was released by removing one or both of the bulky lipophilic riboses and substituting them by linear alkyl chains. The supramolecular behaviour of these new monomers, falling either into the formation of polymers or cyclic species, was already evident from their solubility and gelation ability. Whereas **G1-C$_{alk10}$** exhibited similar characteristics as **G1-C1**, forming gels in a variety of solvents, **G$_{alk2}$-C1** and **G$_{alk10}$-C$_{alk10}$**, although not extraordinarily soluble, were unable to gelate any solvent.

NMR analysis of the different samples further confirmed the self-assembly differences of **G1-C1** – **G1-C$_{alk10}$**, on one hand, and **G$_{alk2}$-C1** – **G$_{alk10}$-C$_{alk10}$**, on the other. Although a wide variety of experiments (concentration- and temperature dependent experiments) in diverse solvents ($CDCl_3$, THF-d_8, $CDCl_2CDCl_2$) were executed, not all conditions provided well-resolved ^{1}H signals. Probably the most informative experiment was the variation of solvent polarity by gradually increasing the volume fraction of DMSO-d_6 in $CDCl_3$-DMSO-d_6 mixtures, which is shown in Figure 5.17 for five different samples: a) a 1:1 mixture of **G** and **C** mononucleosides, b) **G1-C1**, c) **G$_{alk2}$-C1**, d) **G$_{alk10}$-C$_{alk10}$** and e) a longer, previously studied monomer **G1-B1-C1**.

Figure 5.17 reveals the 1:1 mixture of **G** and **C** mononucleosides where throughout all the denaturation process the shape and position of the ^{1}H NMR probe change significantly, suggesting in this case a fast equilibria in the NMR timescale between **G**+**C** complex and **G** and **C** molecules. Due to the reasons commented previously, in this kind of experiments, **G1-C1** (Figure 5.17) shows only the probe of the monomeric G-H^1 amide signal. This behavior suggest the disruption of the gel formation by the addition of increasing amounts of DMSO-d_6 and the subsequent formation of the "*anti*" conformation in solution directed by the steric hindrance,

avoiding thus the cyclotetramerization. **G_{alk2}-C1** offers in Figure 5.17 different set of signals, in which the formation of the cyclic tetramer is patent due to the signals appeared at 13.8 and 9.6 ppm respectively. These broad signals suggests that the cyclic tetramer is not very stable in chloroform and the addition of a few amount of DMSO-d_6 shift the equilibrium towards the monomer formation. However, if both ribose moieties are changed by long alkyl chains, the stability of the cyclic tetramer changes as it is shown in Figure 5.17. Now, **G_{alk10}-C_{alk10}** has the capability to form more stable cyclic tetramers in solution when a few amount of DMSO-d_6 is added. This change is due to reduction of the steric hindrance between N-9 and N-1 substituents so that the population of the "*syn*" conformation is easily reached. Denaturation experiment with increasing amounts of DMSO (*c.a.* 20%) revealed a different scenario, in which the characteristic ^{1}H-NMR signals between tetramer and monomer species and the signals related with oligomeric species formed by a low number of monomers were observed.

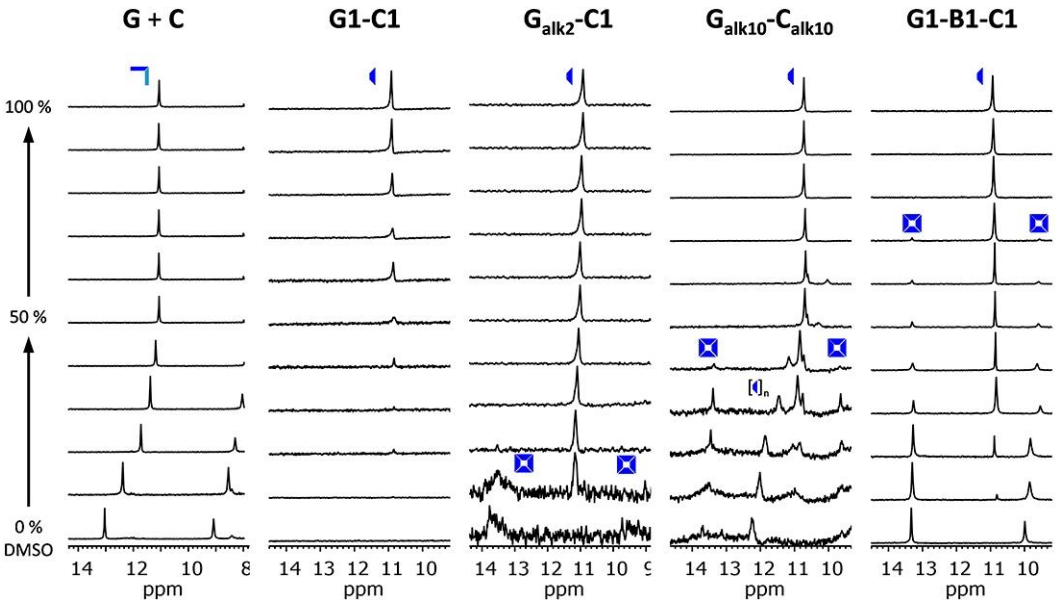


Figure 5.17. Denaturation experiments with increasing amounts of DMSO. Downfield region of the ^{1}H NMR spectra, showing the H-bonded G-amide (H^1; dark blue) and C-amine (H^2; light blue) proton signals, of a 1:1 **G**+**C** mixture and of **G1-C1** – **G1-B1-C1** dinucleosides as the volume fraction of DMSO-d_6 is increased in $CDCl_3$-DMSO-d_6 mixtures at C = 1.0x10^{-2} M and T = 298 K. Additional spectra were recorded at low DMSO content for **G1-C1** – **G_{alk10}-C_{alk10}**, but are not shown here for the sake of clarity and homogeneity.

Steric hindrance.

Steric hindrance has a special key role in the macrocyclization process of this monomers. We can distinguish two different ways in which this factor affects to achieve the "*syn*" conformation that is the only one that can lead the tetramer formation. The first point is the bulky of the substituents on *N*-9 and *N*-1 positions, where **G1-C1** is the best example to explain this effect. In this monomer, we can observe clearly the proximity between ribose moieties in which the steric hindrance is high and thus must be defeated to form the tetrameric species. As we could see previously, this monomer forms gels, suggesting that the "*anti*" conformation is highly populated. In contrast, if both ribose substituents are changed by long alkyl chains (**G_{alk10}-C_{alk10}**), this inconvenient is avoided. Thus, the monomer can rotate

with more freedom reaching the "*syn*" conformation leading in part to the tetrameric formation (Figure 5.18).

The second important reason is the direction in which the N-9 (purine) and N-1 (pyrimidine) substituents are pointing. In this case the purine is pointing to de centre of the molecule whilst the pyrimidine is pointing to the outsides, so that the volume of these moieties and where is placed is important to design the targeted monomer and favour the rotation over the ethynyl group to reach the final "*syn*" conformation.

Once these points are considered, the different tetramerization process between **G1-C_{alk10}** and **G_{alk2}-C1** can be explained. In the first case (**G1-C_{alk10}**), shows the same behavior of **G1-C1**, so that the gelation on low polarity solvents was observed. In contrast, **G_{alk2}-C1** is clearly the opposite and reveals different supramolecular behavior. This molecule is soluble in organic solvents and gel formation was not observed. The main difference is the location of the alkyl chain in the purine, which is by far less bulky than the ribose moiety conferring to the molecule the possibility to spin with less steric problems over the ethynyl group finally reaching the "*syn*" conformation, the only one that afford tetrameric assemblies. Due to this change in the nucleobase substituents, this molecule can form in part cyclic tetramers in solution (Figure 5.18 and 5.19).

As we can observe, the cytidine substituent does not matter if it is bulky or not due to it is pointing to the outsides of the molecule, so that it is not a key factor to the free rotation of the molecule over the triple bond. However, the H proton placed at the C-6 position in the cytidine moiety can offer an additional steric hindrance in the formation of the "*syn*" conformation. This factor will be analysed below in the attempts to synthesize the shortest possible monomer, which the complementary nucleobases are linked through a single carbon-carbon bond.

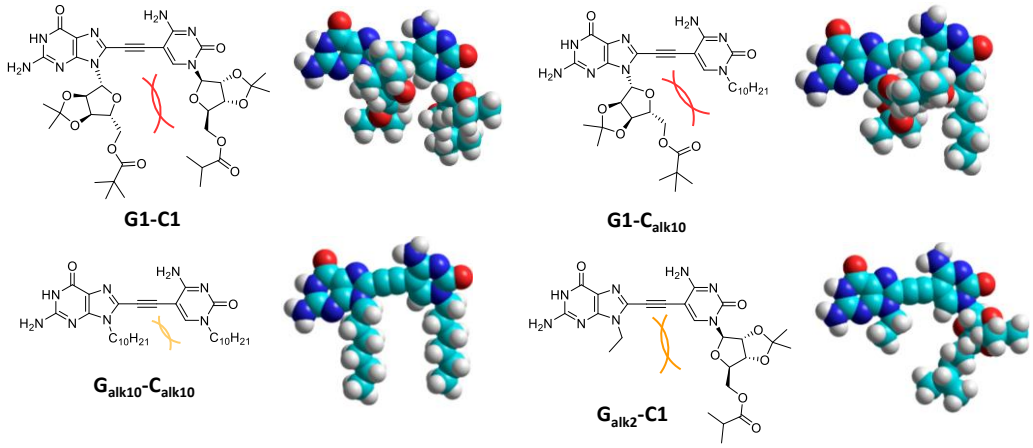


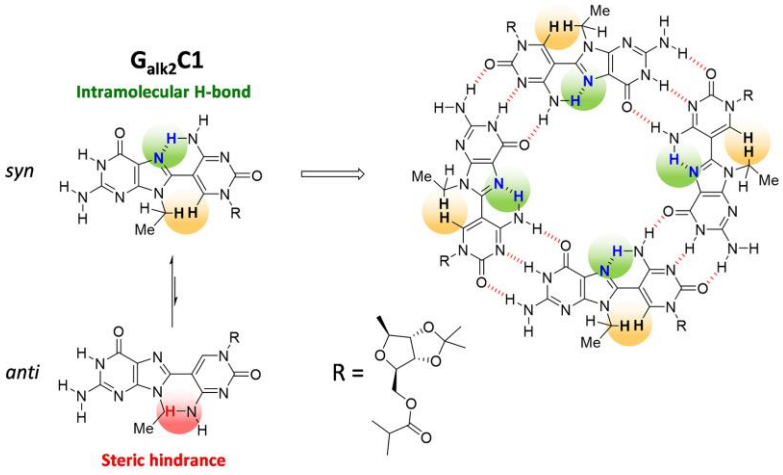
Figure 5.18. Steric hindrance and molecular model (right) of the different ethynyl-linked monomers synthesized in this Chapter. Colour indicate the degree of the steric hindrance, red (strong) and orange (moderate).

5.3.3 Design and Synthetic Attempts toward the shortest monomer.

Concerned by the steric hindrance observed in the ethynyl monomers, molecule **G_{alk2}C1** was also targeted. This molecule, again based on the Watson-Crick GC base-pairing model, is the shortest possible monomer inspired by our common design: its nucleobases are solely

separated by a simple single covalent bond. Besides the absence of the triple bond, monomer **G_{alk2}C1** displays other interesting features. The rotation around this bond is influenced by the fact that the "*syn*" conformation would be favored by the formation of an intramolecular H-bond between H-1 of the C amino group and *N*-7 of the G-nucleobase, leading to a six-membered cycle. In contrast, the "*anti*" conformation would be discouraged by steric reasons (Scheme 5.2): the alkyl chain carried by the G-nucleobase and the C amino group are too close. In these terms, the free rotation of this bond is hindered and we would expect that the "*syn*" conformation should be largely favored compared to the opposite "*anti*" conformation. However, the "*syn*" conformation provokes the steric hindrance between the proton at C-6 in the pyrimidine and the substituent of the purine hindering to adopt a flat disposition although it is not determinant in the cyclization process.

Therefore, molecule **G_{alk2}C1** could potentially form the strongest tetrameric cyclic species, because of the elevated structural predisposition, resulting in high *EM* values.



Scheme 5.2. Chemical structure of the targeted monomer **G_{alk2}C1**, "*syn*" and "*anti*" conformations equilibrium and proposed cyclic tetramer assembly.

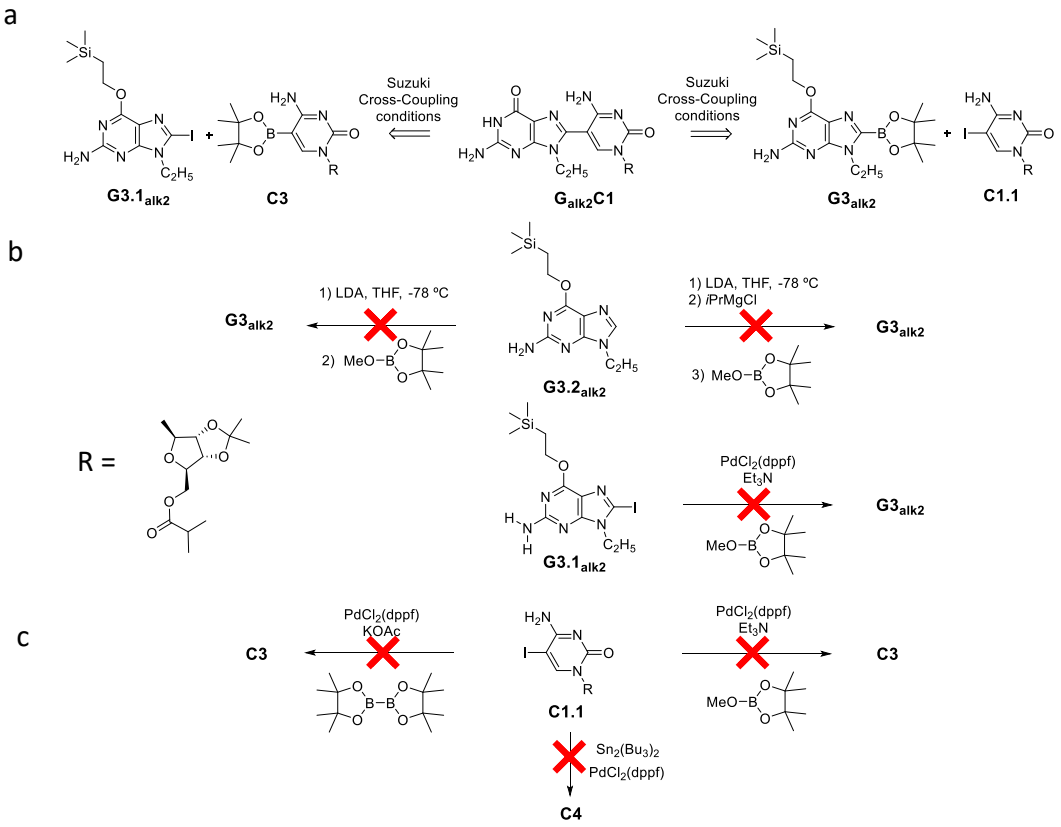
In the synthetic pathway toward **G_{alk2}C1**, different strategies were considered (Scheme 5.3) that relied on Suzuki or Stille cross-couplings. The first retrosynthesis contemplates the coupling between the cytosine derivative **C1** and a borylated-guanine derivative *via* a Suzuki Pd-catalyzed reaction (Scheme 5.3, right). Another possibility would be to incorporate the borane derivative in the C-heterocycle and execute the Suzuki coupling reaction with **G_{alk2}** (Scheme 5.3, left).

For the formation of the boronic esters/acids, deprotonation of the starting product **G3.2$_{alk2}$** with freshly prepared LDA, followed by direct addition of methoxyboronic acid pinacol ester was first essayed (Scheme 5.3b, left).[166] However, the borylation reaction did not occur under these conditions and mostly the starting material **G3.2$_{alk2}$** was recovered. In a second attempt, **G3.2$_{alk2}$** was first deprotonated with LDA; then the Grignard reagent isopropylmagnesium chloride was added; and the resulting organomagnesium-guanine derivative was finally quenched with the dioxoborolane reagent, to afford product **G3$_{alk2}$**

[166] S. J. Mcquaker, C. L. Quinlan, S. T. Caldwell, M. D. Brand, R. C. Hartley, *ChemBioChem.* **2013**, *14*, 993–1000.

(Scheme 5.3, right).[166] Unfortunately, once again only the starting material was recovered from a complex mixture.

In view of these results, a Pd-catalyzed borylation strategy[167] was followed at this point. This protocol was first attempted on **G3.1$_{alk2}$**, using pinacolborane in the presence of NEt_3 and $PdCl_2$(dppf),[168] but only the deiodination product was obtained (Scheme 5.3a, top).



Scheme 5.3. Possible synthetic routes toward **G$_{alk2}$C1**.

Then, we tested the reaction on the aryl iodide **C1.1** both with pinacolborane[168] and bis(pinacolato)diboron in the presence of KOAc and $PdCl_2$(dppf) under strict anhydrous conditions (Scheme 5.3, bottom).[169] Despite these are rather general conditions employed in Pd-catalyzed borylation of haloarenes, none of these attempts afforded the desired borylated product and only the dehalogenated derivatives could be isolated as the major product.

In order to find other synthetic route to reach the desired compound **G$_{alk2}$C1**, the stannilation of the **C1.1** precursor was considered. Unfortunately, the synthesis of the **C4** deriative through the Pd-catalyzed Stille reaction[170] with bis(tributyltin) was fruitless and once

[167] W. Kin Chow, O. Ying Yuen, P. Ying Choy, C. Ming So, C. Po Lau, W. Tak Wong, F. Yee Kwong, *RSC Adv.* **2013**, *3*, 12518–12539.
[168] a) M. Murata, S. Watanabe, Y. Masuda, *J. Org. Chem.* **1997,** *62*, 6458–6459; b) M. Murata, T. Oyama, S. Watanabe, Yuzuru Masuda, **2000**, *65,* 164–168.
[169] T. Ishiyama, M. Murata, N. Miyaura, *J. Org. Chem.* **1995**, *60*, 7508–7510.
[170] J.E. Pickett, A. Váradi, T. C. Palmer, S. G. Grinnell, G. W. Pasternak, R. R. Karimov, S. Majumdar, J. M. Schrock, *Bioorg. Med. Chem. Lett.*, **2015**, *25*, 1761–1764.

again, the starting material or the deiodination derivative were recovered as it is shown in Scheme 5.3c.

5.3.4 The A_{alk10}-U1 monomer.

We also studied the tetramerization process of a given monomer inspired again in our previous model based on Watson-Crick AU base-pairing model (Figure 5.19a). Once the monomer was synthesized according to the previous synthetic routes explained, the solubility problems in $CDCl_3$ afforded a broad ^1H NMR spectra. Then, we monitored the changes observed by ^1H NMR upon increasing DMSO-d_6 content in $CDCl_3$ solutions (Figure 5.19b). **A_{alk10}-U1** shows the apparition of the U-H signal at 11.60 ppm, suggesting that the only species in solution is the monomer since the characteristic H-bonded tetramer signals appear at 13.8 ppm. Is worthy to mention that due to the knowledge acquired in *Chapter 3* in which the symmetry of the multipoint H-bond pattern has a huge impact in the chelate effect, this monomer must show a low tetramer formation that the GC monomers. This predisposition and the low solubility of this compound in apolar organic solvents provoke that the monomer is the only one species present in the ^1H NMR spectra.

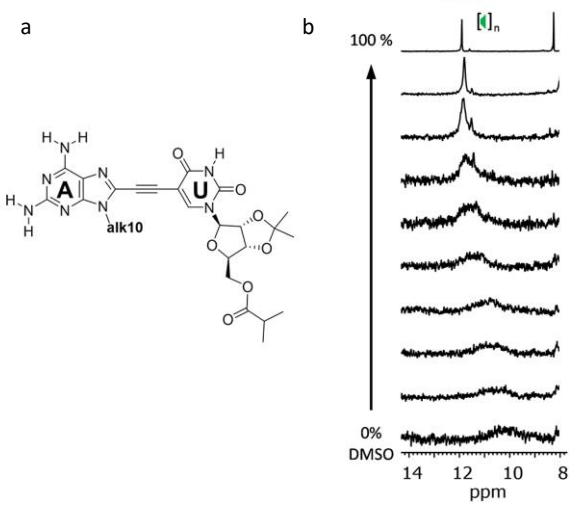


Figure 5.19. (**a**) Chemical structure of **A_{alk10}-U1** and (**b**) downfield region of the ^1H NMR spectra, showing the H-bonded U-imide (H^1; green) proton signals, of **A_{alk10}-U1** dinucleosides as the volume fraction of DMSO-d_6 is increased in $CDCl_3$-DMSO-d_6 mixtures at C = 1.0x10^{-2} M and T = 298 K.

5.4 Conclusions.

In this *Chapter 5* we reported the synthesis of a new family of monomers with different oligophenylene-ethynylene central blocks or triple and single C bonds, based on our previous unsymmetric monomers substituted at both edges by complementary nucleobases. The studies and the analysis of the respective monomers through concentration- and temperature-dependent measurements, as well as with denaturation experiments, clearly demonstrate the impact of each central block placed between complementary nucleobases on the chelate effect in two different ways.

Firstly, when the central block becomes larger, there is a notable loss in the value of the effective molarity due to the bending of the σ-bonds and the free rotation of the oligophenylene-ethynylene central blocks that confers additions degrees of freedom that must be lost upon cyclization. This effect, as we commented previously, is due only to entropic factors.

Secondly, when the central linker becomes shorter as a triple bond or a single C-C bond, other effects become important. Steric hindrance is one of them, and has an important role in order to reach the desired "syn" conformation. The volume and the direction in which the substitution on the nucleobase is pointing is crucial.

Our conclusions could in principle be extended to many supramolecular cycles or cages in which the size is tuned. Thus, the previous design of the respective building blocks and the substituents that confer solubility in organic media must be exquisite in order to reach the desired supramolecular structure quantitatively.

The electronic effects in the non-bonded nucleobase when a Watson-Crick H-bond is formed in the other side of the molecule is being investigated in collaboration with Prof. Dr. Célia Fonseca Guerra, Amsterdam Center for Multiscale Modeling (ACMM) at the University of Leiden.

5.5. Experimental Section.

The General Methods detailed in the Experimental Section **1.4** of *Chapter 1* and Experimental Section **2.3** of *Chapter 2* are also applicable here. Part of the work described in this Chapter can also be found in the Supporting Information of our paper:

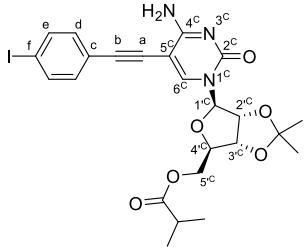
How Large Can we Build a Cyclic Assembly? Impact of Ring Size on Chelate Cooperativity in Noncovalent Macrocyclizations, C. Montoro-García, M. J. Mayoral, R. Chamorro, D. González-Rodríguez.

5.5.1. Synthesis and Characterization.

The synthesis and characterization of compound **G3.1$_{alk10}$**, **G3.2$_{alk10}$**[171] and **G$_{alk10}$-A$_{alk10}$**[121] have been reported elsewhere. Some of the compounds depicted in this Chapter, like **G3.1$_{alk10}$**, **G3.2$_{alk10}$** and **G$_{alk10}$-A$_{alk10}$** were provided from other researchers in our group.

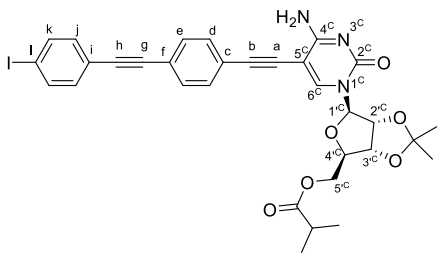
Monocoupling products

Cytidine monocoupling products.



C1-B1-I. C1-B1-I was prepared according to a *Standard Procedure C* between the ethynyl-nucleobase **C1** and 1,4-diiodobenzene. A dry a THF/NEt_3 (4:1) mixture (8 mL) was poured over **C1** (566 mg, 1.5 mmol), 1,4-diiodobenzene (3.5 g 10.5 mmol), $PdCl_2(PPh_3)_2$ (9.5 mg, 0.01 mmol), and CuI (1.3 mg, 7 µmol). The mixture was stirred under argon for 12 h at 40 ºC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording **C1-B1-I** as a yellow solid (730 mg, 84 %). The excess of 1,4-diiodobenzene was recovered.

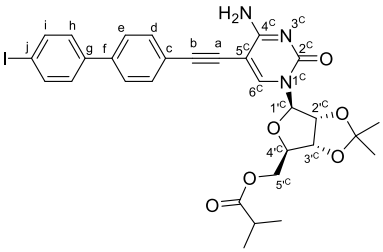
^{1}H RMN (300 MHz, $CDCl_3$) δ(ppm) = 8.57 (s, 1H, NH^{4C}), 7.67 (d, J = 4.1 Hz, 2H, H^{d}), 7.63 (s, 1H,H^{6C}), 7.12 (d, J = 8.4 Hz, 2H, H^{e}), 5.83 (s, 1H, NH^{4C}), 5.65 (d, J = 1.7 Hz, 1H, $H^{1'C}$), 4.90 (dd, J = 6.3, J' = 1.7 Hz, 1H, $H^{2'C}$), 4.73 (dd, J = 6.3, J' = 3.7 Hz, 1H, $H^{3'C}$), 4.47 – 4.14 (m, 3H, $H^{4'C}$, $H^{5'C}$), 2.50 (hep, 1H, J = 7.0 Hz, -OCOCH(CH_3)$_2$), 1.51 (s, 3H, -OC(CH_3)), 1.28 (s, 3H, -OC(CH_3)), 1.08 (dd, J = 7.0, J' = 2.5 Hz, 6H, -COCH-(CH_3)$_2$).



C1-B2-I. C1-B2-I was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between the ethynyl-nucleobase **C1** and **B2**. A dry a THF/NEt_3 (4:1) mixture (10 mL) was poured over **B2** (225 mg, 0.52 mmol), $PdCl_2(PPh_3)_2$ (1.83 mg, 2.01 µmol), and CuI (0.25 mg, 1.30 µmol). Then, **C** (50 mg, 0.13 mmol) was added dropwise and the mixture was stirred under argon for 12h at 40 ºC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (60:1), affording **C1-B2-I** as a brown solid (117 mg, 88 %). The excess of **B2** was recovered.

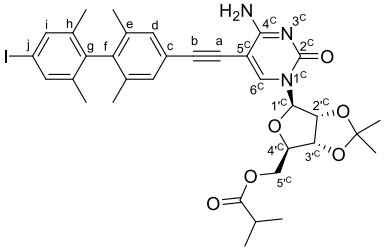
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.07 (s (broad), 1H, NH^{4C}), 7.73 (s, 1H, H^{6C}), 7.68 – 7.58 (m, 4H, $H^{k,j}$), 7.24 – 7.09 (m, 4H, $H^{e,d}$), 6.12 (s (broad), 1H, NH^{4C}), 5.69 (s, 1H, $H^{1'C}$), 4.91 (d, J = 6.2 Hz, 1H, $H^{2'C}$), 4.74 (dd, J = 6.5, J' = 3.5 Hz, 1H, $H^{3'C}$), 4.42 – 4.14 (m, 3H, $H^{4'C}$, $H^{5'C}$), 2.49 (h, J = 7.3 Hz, 1H, -COCH), 1.49 (s, 3H, -OC(CH_3)), 1.27 (s, 3H, -OC(CH_3)), 1.07 (dd, J = 7.0, J' = 2.9 Hz, 6H, -COCH-(CH_3)$_2$).

[171] N. Bilbao, V. Vázquez-González, M. T. Aranda, D. González-Rodríguez, *Eur. J. Org. Chem.* **2015**, *54*, 7160–7175.



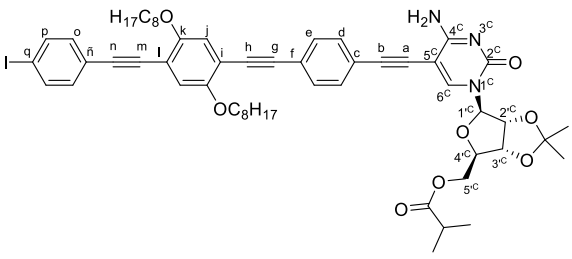
C1-B2'-I. C1-B2'-I was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between the ethynyl-nucleobase **C1** and **B2'**. A dry a THF/NEt_3 (4:1) mixture (10 mL) was poured over **B2'** (537 mg, 1.32 mmol), $PdCl_2(PPh_3)_2$ (3.71 mg, 5.2 µmol), and CuI (0.5 mg, 2.65 µmol). Then, **C1** (100 mg, 0.27 mmol) was added dropwise and the mixture was stirred under argon for 12h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1), affording **C1-B2-I** as a brown solid (120 mg, 67 %). The excess of **B2'** was recovered.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.07 (s (broad), 1H, NH^{4C}), 7.76 (s, 1H, H^{6C}), 7.70 (d, J = 8.2 Hz, 2H, H^{i}), 7.48 (s, 4H, $H^{h,e}$), 7.28 (d, J = 8.2 Hz, 2H, H^{e}), 6.19 (s (broad), 1H, NH^{4C}), 5.70 (s, 1H, $H^{1'C}$), 4.91 (d, J = 6.5 Hz, 1H, $H^{2'C}$), 4.75 (t, J = 5.1 Hz, 1H, $H^{3'C}$), 4.37 – 4.16 (m, 3H, $H^{4'C}$, $H^{5'C}$), 2.51 (hept, J = 6.9 Hz, 1H, -COCH), 1.49 (s, 3H, -OC(CH_3)), 1.27 (s, 3H, -OC(CH_3)), 1.08 (d, J = 7.0 Hz, 6H, -COCH-(CH_3)$_2$).



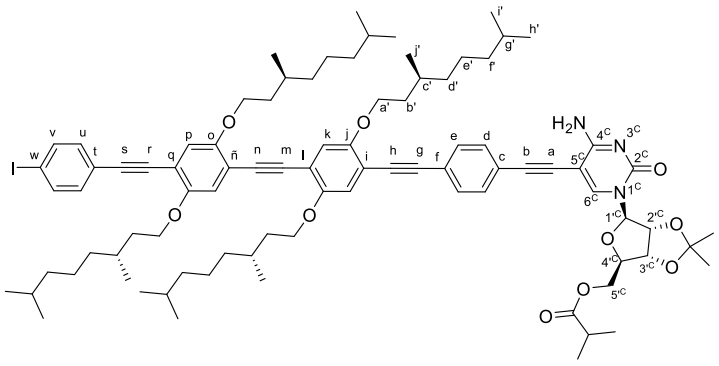
C1-B2''-I. C1-B2''-I was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between the ethynyl-nucleobase **C1** and **B2''**. A dry a THF/NEt_3 (4:1) mixture (13 mL) was poured over **B2''** (322 mg, 0.7 mmol), $PdCl_2(PPh_3)_2$ (0.19 mg, 0.2 µmol), and CuI (0.02 mg, 0.13 µmol). Then, **C1** (52 mg, 13 µmol) was added dropwise and the mixture was stirred under argon for 12h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1), affording **C1-B2-I** as a brown solid (70 mg, 71 %). The excess of **B2''** was recovered.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.90 (s (broad), 1H, NH^{4C}), 7.66 (s, 1H, H^{6C}), 7.43 (s, 2H, H^{i}), 5.92 (s, 2H, H^{d}), 5.67 (d, J = 3.2 Hz, 1H, $H^{1'C}$), 4.92 (d, J = 6.2 Hz, 1H, $H^{2'C}$), 4.74 (dd, J = 6.3, J' = 3.6 Hz, 1H, $H^{3'C}$), 4.44 – 4.18 (m, 3H, $H^{4'C}$, $H^{5'C}$), 2.53 (pd, J = 7.0, J' = 2.1 Hz, 1H, -COCH), 1.91 – 1.73 (m, 12H, -CCH_3), 1.50 (s, 3H, -OC(CH_3)), 1.28 (s, 3H, -OC(CH_3)), 1.09 (dd, J = 7.0, J' = 2.3 Hz, 6H, -COCH-(CH_3)$_2$).



C1-B3-I. C1-B3-I was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between the ethynyl-nucleobase **C1** and **B3**. A dry a THF/NEt_3 (4:1) mixture (15 mL) was poured over **B3** (1.04 g, 1.32 mmol), $PdCl_2(PPh_3)_2$ (3.71 mg, 5.29 µmol), and CuI (0.50 mg, 2.65 µmol). Then, **C1** (100 mg, 0.27mmol) was added dropwise and the mixture was stirred under argon for 12h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (40:1), affording **C1-B3-I** as a brown solid (190 mg, 69 %). The excess of **B3** was recovered.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.33 (s (broad), 1H, NH^{4C}), 7.76 (s, 1H, H^{6C}), 7.69 (d, J = 8.4 Hz, 2H, H^{p}), 7.52 (d, J = 8.3 Hz, 2H, H^{o}), 7.44 (d, J = 8.2 Hz, 2H, H^{e}), 7.25 (d, J = 7.8 Hz, 2H, H^{d}), 7.00 (d, J = 2.0 Hz, 2H, H^{j}), 5.89 (s, 1H, NH^{4C}), 5.73 (s, 1H, $H^{1'C}$), 4.98 (dd, J = 6.3, J' = 1.6 Hz, 2H, $H^{2'C}$), 4.81 (dd, J = 6.3, J' = 3.7 Hz, 1H, $H^{3'C}$), 4.50 – 4.27 (m, 3H, $H^{4'C}$, $H^{5'C}$), 4.03 (td, J = 6.5, J' = 2.1 Hz, 4H, -(OCH_2)-), 2.58 (p, J = 7.0 Hz, 1H, -COCH), 1.94 – 1.77 (m, 4H, -OCH$_2$CH_2-), 1.58 (s, 3H, -OC(CH_3)), 1.35 (s, 3H, -OC(CH_3)), 1.30 – 1.25 (m, 24H, -CH_2-), 1.16 (dd, J = 7.0, J' = 2.6 Hz, 6H, -COCH-(CH_3)$_2$), 0.86 (t, 6H, (-CH$_2$CH_3)$_2$).

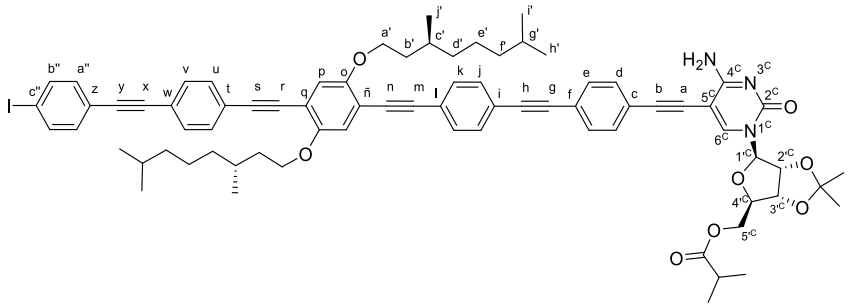


C1-B4-I. C1-B4-I was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between the ethynyl-nucleobase **C1** and **B4**. A dry a THF/NEt_3 (4:1) mixture (10 mL) was poured over **B4** (124 mg, 0.1 mmol), $PdCl_2(PPh_3)_2$ (0.69 mg, 1 µmol), and CuI (0.09 mg, 0.5 µmol). Then, **C1** (19 mg, 0.05 mmol) was added dropwise and the mixture was stirred under argon for 12h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (40:1), affording **C1-B4-I** as a brown solid (45 mg, 61 %). The excess of **B4** was recovered.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.63 (s (broad), 1H, N*H*^{4C}), 7.76 (s, 1H, *H*^{6C}), 7.69 (d, *J* = 7.9 Hz, 2H, *H*^{d}), 7.58 – 7.40 (m, 4H, *H*^{e,u}), 7.25 (d, *J* = 8.3 Hz, 2H, *H*^{v}), 7.02 (d, *J* = 5.6 Hz, 4H, *H*^{k,j,o,p}), 5.94 (s, 1H, N*H*^{4C}), 5.74 (s, 1H, *H*$^{1'C}$), 5.00 (d, *J* = 6.2 Hz, 1H, *H*$^{2'C}$), 4.82 (dd, *J* = 6.3, *J'* = 3.4 Hz, 1H, *H*$^{3'C}$), 4.53 – 4.27 (m, 3H, *H*$^{4'C}$, *H*$^{5'C}$), 4.07 (q, *J* = 9.2, *J'* = 8.3 Hz, 8H, -OC*H*$_2$), 2.59 (hept, *J* = 7.1 Hz, 1H, -COC*H*), 2.01 – 1.04 (m, 52H, -C*H*$_2$$^{b'-g'}$-, -OC(C*H*$_3$), -COCH-(C*H*$_3$)$_2$), 0.99 (d, *J* = 5.8 Hz, 12H, -C*H*$_3$$^{j'}$), 0.85 (d, *J* = 6.7 Hz, 24H, -C*H*$_3$$^{i',h'}$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 176.5, 164.8, 153.8, 153.7, 153.7, 153.5, 145.1, 137.5, 133.0, 132.2, 132.0, 131.6, 131.4, 129.0, 128.6, 128.4, 128.2, 125.3, 124.2, 123.0, 121.5, 117.0, 116.9, 114.6, 114.5, 114.2, 113.5, 113.5, 95.7, 95.3, 94.3, 94.1, 93.9, 91.7, 91.6, 91.4, 88.6, 87.5, 85.8, 85.6, 81.2, 81.0, 68.2, 67.9, 67.8, 64.2, 39.3, 37.4, 37.3, 36.3, 33.9, 30.1, 30.0, 30.0, 29.7, 28.0, 27.2, 25.3, 24.8, 24.8, 24.7, 22.7, 22.6, 19.9, 19.8, 19.7, 19.1, 18.9.

HRMS (ESI+): Calculated for $C_{88}H_{118}IN_3O_{10}$: 1505.8275 [M+H]$^+$. Found: 1526.7714 [M+Na]$^+$**.**



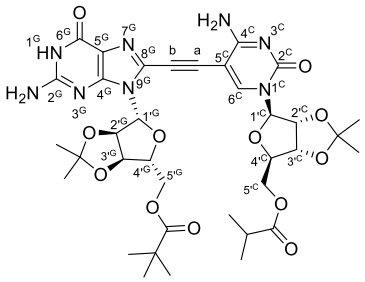
C1-B5-I. C1-B5-I was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between the ethynyl-nucleobase **C1** and **B5**. A dry a THF/NEt_3 (4:1) mixture (10 mL) was poured over **B5** (0.27 g, 0.26 mmol), $PdCl_2(PPh_3)_2$ (0.73 mg, 1 µmol), and CuI (0.09 mg, 0.5 µmol). Then, **C** (20 mg, 0.05 mmol) was added dropwise and the mixture was stirred under argon for 12h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (40:1), affording **C1-B5-I** as a brown solid (49 mg, 73 %). The excess of **B5** was recovered.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.95 (s (broad), 1H, N*H*^{4C}), 7.77 (s, 1H, *H*^{6C}), 7.70 (d, *J* = 8.3 Hz, 2H, *H*^{d}), 7.51 (d, *J* = 4.7 Hz, 12H, *H*$^{e,j,k,u,v,a''}$), 7.25 (d, *J* = 8.1 Hz, 2H, *H*$^{b''}$), 7.02 (s, 2H, *H*^{p,o}), 5.87 (s, 1H, N*H*^{4C}), 5.74 (d, *J* = 1.9 Hz, 1H, *H*$^{1'C}$), 4.99 (d, *J* = 6.6 Hz, 1H, *H*$^{2'C}$), 4.82 (dd, *J* = 6.3, *J'* = 3.6 Hz, 1H, *H*$^{3'C}$), 4.49 – 4.28 (m, 3H, *H*$^{4'C}$, *H*$^{5'C}$), 4.17 – 3.99 (m, 4H, -(OC*H*$_2$)-), 2.66 – 2.51 (m, 1H, -COC*H*), 1.99 – 1.09 (m, 32H, -C*H*$_2$$^{b'-g'}$-, -OC(C*H*$_3$), -COCH-(C*H*$_3$)$_2$), 0.98 (d, *J* = 6.5 Hz, 6H, -C*H*$_3$$^{j'}$), 0.86 (d, *J* = 6.6 Hz, 12H, -C*H*$_3$$^{i',h'}$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 176.5, 164.7, 153.7, 145.2, 141.3, 137.6, 133.1, 132.6, 132.4, 131.7, 131.6, 131.5, 131.4, 123.7, 123.6, 122.7, 122.7, 122.6, 116.8, 114.2, 113.9, 95.7, 94.6, 94.4, 91.6, 90.6, 88.1, 85.8, 85.6, 81.0, 68.0, 64.2, 39.3, 37.4, 36.3, 33.9, 30.9, 30.0, 29.7, 28.0, 27.2, 25.3, 24.8, 22.7, 22.6, 19.7, 19.1, 18.9.

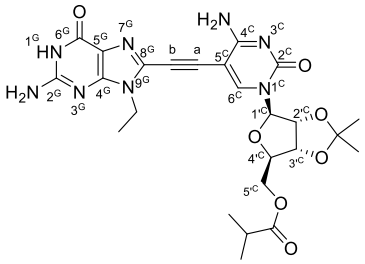
HRMS (ESI+): Calculated for $C_{76}H_{82}IN_3O_8$: 1292.5225 [M+H]$^+$. Found: 1314.5046 [M+Na]$^+$.

Monomers



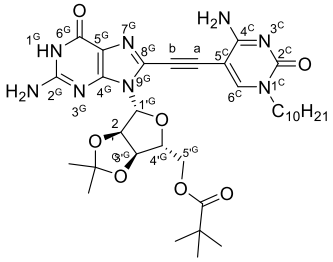
G1-C1. G1-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between iodo-nucleobase **C1.1** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (10 mL) was poured over **G1** (200 mg 0.5 mmol), $PdCl_2(PPh_3)_2$ (6.5 mg, 9.3 µmol), and CuI (0.88 mg, 4.6 µmol). Then, **C1.1** (289 mg, 0.6 mmol) was added and the mixture was stirred under argon for 12h at 40 °C. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by recrystallization in acetonitrile, affording **G1-C1** as a white solid (316 mg, 81 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.00 (s (broad), 1H, NH^{1G}), 8.28 (s, 1H, H^{6C}), 8.00 (s (broad), 1H, NH^{4C}), 7.37 (s (broad), 1H, NH^{4C}), 6.80 (s (broad), 2H, N$H_2{}^{2G}$), 6.13 (s, 1H, $H^{1'G}$), 5.81 (s, 1H, $H^{1'C}$), 5.43 (d, J = 6.3 Hz, 1H, $H^{2'G}$), 5.22 (dd, J = 6.2, J' = 4.0 Hz, 1H, $H^{2'C}$), 5.02 (dd, J = 6.4, J' = 1.6 Hz, 1H, $H^{3'G}$), 4.91 – 4.76 (m, 1H, $H^{3'C}$), 4.42 – 3.97 (m, 6H, $H^{4'G}$, $H^{5'G}$, $H^{4'C}$, $H^{5'C}$), 2.61 – 2.54 (m, 1H, COCH), 1.50 (d, J = 7.9 Hz, 6H, -OC(CH_3)), 1.31 (d, J = 9.8 Hz, 6H, -OC(CH_3)), 1.08 (d, J = 10.3 Hz, 14H, -COCH-(CH_3)$_2$).
^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 179.4, 177.8, 164.6, 159.5, 155.7, 154.4, 150.8, 145.0, 121.8, 118.1, 114.2, 113.4, 95.8, 94.6, 93.2, 92.4, 89.3, 87.7, 85.4, 85.0, 84.2, 83.5, 81.4, 79.4, 79.1, 77.8, 76.5, 76.4, 76.1, 66.3, 64.2, 38.1, 33.2, 29.4, 27.8, 27.6, 27.5, 25.8, 25.6, 19.5.
HRMS (FAB+): Calculated for $C_{36}H_{47}N_8O_{12}$: 783.3313 $[M+H]^+$. Found: 783.3334 $[M+H]^+$**.**



G_{alk2}-C1. G_{alk2}-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between iodo-nucleobase **C1.1** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (10 mL) was poured over, $\mathbf{G_{alk2}}$[149] (60 mg 0.3 mmol), $PdCl_2(PPh_3)_2$ (4.1 mg, 6 µmol), and CuI (0.6 mg, 0.3 µmol). Then, **C1.1** (283 mg, 0.6 mmol), was added and the mixture was stirred under argon for 12h at 40 °C. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by recrystallization in acetonitrile, affording $\mathbf{G_{alk2}}$**-C1** as an orange solid (123 mg, 75 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.90 (s (broad), 1H, NH^{1G}), 8.26 (s, 1H, H^{6C}), 7.96 (s (broad), 1H, NH^{4C}), 7.26 (s (broad), 1H, NH^{4C}), 6.82 (s (broad), 2H, N$H_2{}^{2G}$), 5.79 (s, 1H, $H^{1'C}$), 5.03 (d, J = 6.3 Hz, 1H, $H^{2'C}$), 4.82 (d, J = 6.3 Hz, 1H, $H^{3'C}$), 4.26 (d, J = 7.9 Hz, 3H, $H^{4'C}$, $H^{5'C}$), 4.07 (q, J = 7.1, 6.5 Hz, 2H, -CH_2CH$_3$), 2.64 – 2.52 (m, 1H, COCH), 1.48 (s, 3H, -OC(CH_3)), 1.31 (d, J = 7.3 Hz, 6H, -OC(CH_3), -CH$_2$CH_3), 1.06 (dd, J = 7.0, J' = 4.3 Hz, 6H, -COCH-(CH_3)$_2$).
^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 175.8, 164.0, 156.1, 154.4, 152.9, 150.8, 148.8, 129.4, 117.0, 112.9, 94.1, 88.4, 85.0, 84.4, 84.2, 83.7, 80.9, 64.0, 57.5, 57.5, 37.8, 33.1, 26.9, 25.1, 23.0, 19.2, 18.7, 18.6, 14.8, 13.4.
HRMS (ESI+): Calculated for $C_{25}H_{31}N_8O_7$: 555.2316 $[M+H]^+$. Found: 555.2307 $[M+H]^+$**.**

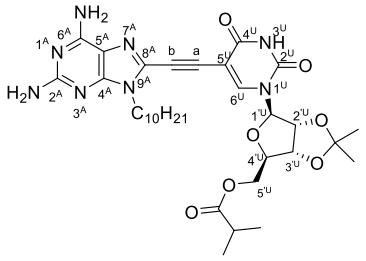


G1-C_{alk10}. G1-C_{alk10} was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between iodo-nucleobase $\mathbf{C_{alk10}}$[27] and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (12 mL) was poured over $\mathbf{C_{alk10}}$ (100 mg 0.23 mmol), $PdCl_2(PPh_3)_2$ (3.3 mg, 4.6 µmol), and CuI (0.4 mg, 2.3 µmol). Then, **G1** (105 mg, 0.28 mmol), was added and the mixture was stirred under argon for 12h at 40 °C. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (10:1), affording $\mathbf{G1\text{-}C_{alk10}}$ as a white solid (149 mg, 73 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.97 (s (broad), 1H, NH^{1G}), 8.31 (s, 1H, H^{6C}), 7.69 (s (broad), 1H, NH^{4C}), 7.09 (s (broad), 1H, NH^{4C}), 6.77 (s (broad), 2H, NH^{2G}), 6.13 (s, 1H, $H^{1'G}$), 5.45 (d, J = 6.2 Hz, 1H, $H^{2'G}$), 5.19 (dd, J = 6.4, J' = 3.4 Hz, 1H, $H^{3'G}$), 4.29 – 4.03 (m, 3H, $H^{4'G}$, $H^{5'G}$), 3.71 (t, J = 7.1 Hz, 2H, -NCH_2-), 1.51 (s, 3H, -OC(CH_3)), 1.33 (s, 3H, -OC(CH_3)), 1.24 (s, 16H, -CH_2-), 1.09 (s, 9H, -COC(CH_3)$_3$), 0.91 – 0.79 (m, 3H, -CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 177.6, 164.6, 156.7, 154.6, 154.3, 152.4, 150.7, 129.9, 117.5, 113.7, 88.9, 87.0, 86.3, 85.8, 84.0, 83.8, 82.0, 64.7, 49.7, 38.6, 31.7, 29.4, 29.3, 29.1, 29.1, 29.0, 27.4, 27.2, 26.3, 25.7, 22.5, 14.4, 0.6.

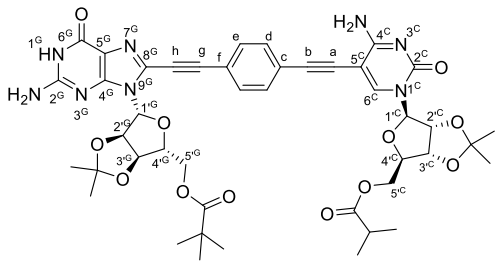
HRMS (MALDI): Calculated for $C_{34}H_{49}N_8O_7$: 681.3719 $[M+H]^+$. Found: 681.3740 $[M+H]^+$. Matrix: DCTB



A_{alk10}-U1. A_{alk10}-U1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between bromo-nucleobase **A_{alk10}**[27] and the ethynyl-nucleobase **U1**. A dry a THF/NEt_3 (4:1) mixture (12 mL) was poured over **A_{alk10}** (50 mg 1.4 mmol), $PdCl_2(PPh_3)_2$ (1.9 mg, 3 µmol), and CuI (0.26 mg, 1.4 µmol). Then, **U1** (77 mg, 2 mmol), was added and the mixture was stirred under argon for 12h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (30:1), affording **A_{alk10}-U1** as a white solid (48 mg, 51 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 7.73 (s (broad), 1H, NH^{3U}), 5.70 (s (broad), 2H, NH^{2A}), 4.93 (s, 1H, $H^{1'U}$), 4.72 (s, 1H, $H^{2'U}$), 4.33 – 4.07 (m, 8H, $H^{3'U}$, $H^{4'U}$, NH^{2A}, -NCH_2-), 4.07 (s, 2H, $H^{5'U}$), 2.55 (d, J = 7.0 Hz, 1H, -COCH), 1.52 (s, 3H, -OC(CH_3)), 1.30 (s, 3H, -OC(CH_3)), 1.22 – 1.06 (m, 19H, -COCH-(CH_3)$_2$, -CH_2-), 0.77 (s, 3H, -CH_3).

HRMS (ESI+): Calculated for $C_{33}H_{47}N_8O_7$: 667.3562 $[M+H]^+$. Found: 667.3584$[M+H]^+$.



G1-B1-C1. G1-B1-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between **C1-B1-I** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (10 mL) was poured over **C1-B1-I** (136 mg, 0.2 mmol), $PdCl_2(PPh_3)_2$ (5.3 mg, 5 µmol), and CuI (0.4 mg, 2 µmol). Then, **G1** (116 mg, 0.3 mmol), was added dropwise and the mixture was stirred under argon for 12h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by recrystallization in acetonitrile, affording **G1-B1-C1** as a yellow solid (110 mg, 54 %).

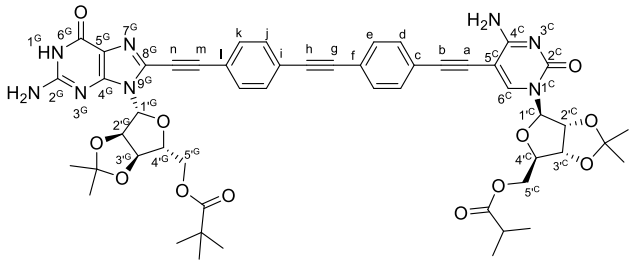
^{1}H RMN (300 MHz, DMSO-d_6) δ(ppm) = 11.00 (s (broad), 1H, NH^{1G}), 8.15 (s, 1H, H^{6C}), 8.00 (s (broad), 1H, NH^{4C}), 7.69 (d, J = 1.1 Hz, 4H, $H^{c,d}$), 7.30 (s (broad), 2H, NH_2^{2G}), 6.79 (s (broad), 1H, NH^{4C}), 6.13 (d, J = 1.4 Hz, 1H, $H^{1'G}$), 5.83 (d, J = 1.8 Hz, 1H, $H^{1'C}$), 5.48 – 5.39 (m, 1H, $H^{2'G}$), 5.28 (dd, J = 6.4, J' = 3.3 Hz, 1H, $H^{2'C}$), 5.02 (dd, J = 6.3, J' = 1.8 Hz, 1H, $H^{3'G}$), 4.83 (dd, J = 6.6, J' = 3.1 Hz, 1H, $H^{3'C}$), 4.39 – 4.05 (m, 6H, $H^{4'G}$, $H^{5'G}$, $H^{4'C}$, $H^{5'C}$), 2.68 – 2.53 (m, 1H, COCH), 1.52 (d, J = 14.2 Hz, 6H, -OC(CH_3)), 1.32 (d, J = 12.0 Hz, 6H, -OC(CH_3)), 1.10 (d, J = 2.3 Hz, 9H, -C(CH_3)$_3$), 1.08 (dd, J = 7.0 Hz, J' = 2.3 Hz, 6H, -COCH-(CH_3)$_2$).

^{13}C-NMR (75 MHz, $CDCl_3$) δ(ppm) = 179.8, 177.4, 164.4, 159.1, 155.00, 154.0, 150.2, 145.4, 131.7, 131.5, 131.0, 121.2, 118.6, 114.7, 113.7, 95.9, 94.2, 93.9, 92.7, 89.8, 87.5, 85.3, 85.1, 84.6, 83.0, 81.2, 79.7, 79.4, 77.2, 76.8, 76.3, 76.3, 66.9, 64.8, 38.9, 33.8, 29.7, 27.3, 27.2, 27.1, 25.4, 25.0, 19.00.

HRMS (ESI+): Calculated for $C_{44}H_{51}N_8O_{12}$: 883.3548 $[M+H]^+$. Found: 883.3657 $[M+H]^+$**.**

UV-vis: λ_{max} (monomer, DMAC, 298 K) = 360 nm (ε = 43703 $M^{-1}cm^{-1}$), 387 nm (sh).

Emission: λ_{max} (monomer, DMAC, 298 K) = 463 nm.



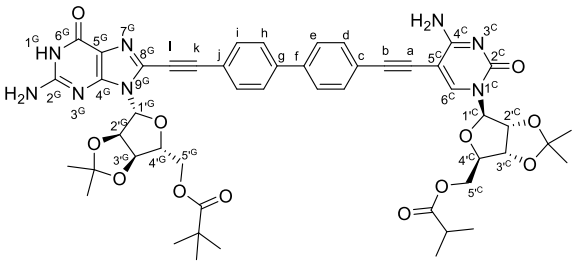
G1-B2-C1. G1-B2-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between **C1-B2-I** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (12 mL) was poured over **C1-B2-I** (78.7 mg, 0.12 mmol), $PdCl_2(PPh_3)_2$ (1.61 mg, 2.3 µmol), and CuI (0.22 mg, 1.15 µmol). Then, **G1** (59.4 mg, 0.14 mmol), was added dropwise and the mixture was stirred under argon for 12h at 40 ^oC. Once completed, the mixture was filtrated over a celite plug and the

solvent was evaporated under reduced pressure. The product was purified by recrystallization in acetonitrile, affording **G1-B2-C1** as a yellow solid (70 mg, 72 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.97 (s (broad), 1H, NH^{1G}), 8.13 (s, 1H, H^{6C}), 7.97 (s (broad), 1H, NH^{4C}), 7.77 – 7.58 (m, 8H, $H^{d,e,j,k}$), 7.26 (s (broad), 1H, NH^{4C}), 6.77 (s (broad), 2H, N$H_2{}^{2G}$), 6.13 (d, J = 1.4 Hz, 1H, $H^{1'G}$), 5.82 (d, J = 1.8 Hz, 1H, $H^{1'C}$), 5.45 (dd, J = 6.2, J' = 1.4 Hz, 1H, $H^{2'G}$), 5.28 (dd, J = 6.4, J' = 3.4 Hz, 1H, $H^{2'C}$), 5.02 (dd, J = 6.4, J' = 1.8 Hz, 1H, $H^{3'G}$), 4.82 (dd, J = 6.3, J' = 3.2 Hz, 1H, $H^{3'C}$), 4.37 – 4.09 (m, 6H, $H^{4'G}$, $H^{5'G}$, $H^{4'C}$, $H^{5'C}$), 2.58 (p, J = 7.1 Hz, 1H, -COCH), 1.52 (d, J = 15.4 Hz, 6H, -OC(CH_3)$_2$), 1.32 (d, J = 13.0 Hz, 6H, -OC(CH_3)$_2$), 1.09 (s, 9H, -C(CH_3)$_3$), 1.07 (d, J = 2.5 Hz, 3H, -COCH-(CH_3)$_2$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 179.7, 177.4, 170.6, 164.6, 159.2, 155.1, 154.0, 150.2, 137.6, 131.7, 131.5, 131.4, 126.9, 114.7, 113.8, 96.0, 94.1, 92.8, 91.9, 90.8, 90.1, 87.3, 85.3, 85.1, 84.7, 83.1, 81.3, 79.3, 64.8, 38.9, 33.9, 31.9, 31.5, 30.2, 29.7, 29.7, 29.4, 27.3, 27.2, 27.1, 25.4, 25.2, 22.7, 19.0, 14.1, 1.0.

HRMS (ESI+): Calculated for $C_{52}H_{54}N_8O_{12}$: 983.3939 [M+H]$^+$. Found: 1005.3765 [M+Na]$^+$**.**

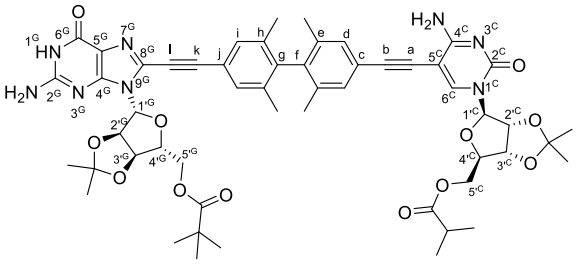


G1-B2'-C1. G1-B2'-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between **C1-B2'-I** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (12 mL) was poured over **C1-B2'-I** (117 mg, 0.18 mmol), $PdCl_2(PPh_3)_2$ (2.5 mg, 3.6 µmol), and CuI (0.03 mg, 1.78 µmol). Then, **G1** (93 mg, 0.21 mmol), was added dropwise and the mixture was stirred under argon for 12h at 40 ^{o}C. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by recrystallization in methanol, affording **G1-B2'-C1** as a yellow solid (75 mg, 45 %).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 10.97 (s (broad), 1H, NH^{1G}), 8.12 (s, 1H, H^{6C}), 7.97 (s (broad), 1H, NH^{4C}), 7.87 (d, J = 8.3 Hz, 2H, H^{d}), 7.82 (d, J = 8.4 Hz, 2H, H^{i}), 7.73 (t, J = 7.8 Hz, 4H, $H^{h,e}$), 7.24 (s (broad), 1H, NH^{4C}), 6.76 (s (broad), 2H, N$H_2{}^{2G}$), 6.15 (s, 1H, $H^{1'G}$), 5.83 (s, 1H, $H^{1'C}$), 5.45 (d, J = 6.2 Hz, 1H, $H^{2'G}$), 5.25 (dt, J = 14.3, J' = 5.1 Hz, 1H, $H^{2'C}$), 5.02 (d, J = 6.4 Hz, 1H, $H^{3'G}$), 4.89 – 4.78 (m, 1H, $H^{3'C}$), 4.41 – 4.06 (m, 6H, $H^{4'G}$, $H^{5'G}$, $H^{4'C}$, $H^{5'C}$), 2.57 (dd, J = 14.2, J' =7.2 Hz, 1H, -COCH), 1.52 (d, J = 15.5 Hz, 6H, -OC(CH_3)$_2$), 1.32 (d, J = 13.3 Hz, 6H, -OC(CH_3)$_2$), 1.10 (d, J = 2.3 Hz, 15H, -C(CH_3)$_3$, -COCH-(CH_3)$_2$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 177.1, 175.8, 156.1, 154.6, 154.2, 150.3, 146.9, 140.3, 138.6, 132.3, 131.9, 127.1, 126.7, 122.1, 119.5, 117.5, 113.3, 113.0, 93.8, 93.1, 88.8, 85.4, 84.6, 84.4, 83.4, 81.5, 80.7, 79.6, 64.2, 63.9, 38.1, 33.1, 27.0, 26.9, 26.8, 25.3, 25.3, 25.1, 18.8, 18.7.

HRMS (ESI+): Calculated for $C_{50}H_{55}N_8O_{12}$: 959.3939 [M+H]$^+$. Found: 959.3974 [M+H]$^+$.

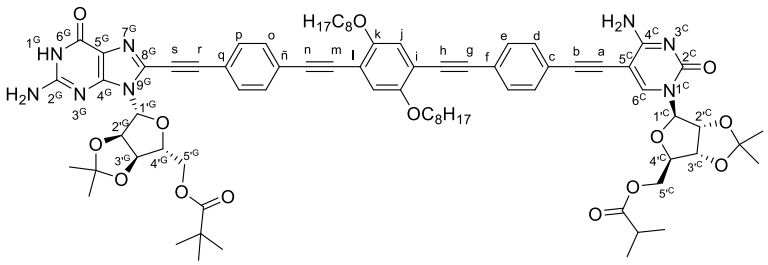


G1-B2''-C1. G1-B2''-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between **C1-B2''-I** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (12 mL) was poured over **C1-B2''-I** (70 mg, 0.1 mmol), $PdCl_2(PPh_3)_2$ (1.4 mg, 2 µmol), and CuI (0.19 mg, 0.1 µmol). Then, **G1** (51 mg, 0.18 mmol), was added dropwise and the mixture was stirred under argon for 12h at 40 ^{o}C. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was purified by recrystallization in methanol, affording **G1-B2''-C1** as an orange solid (40 mg, 70 %).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 11.42 (s (broad), 1H, NH^{1G}), 8.14 (s (broad), 1H, NH^{4C}), 7.84 (s, 1H, H^{6C}), 7.27 (d, J = 13.8 Hz, 4H, $H^{d,i}$), 6.60 (s (broad), 1H, NH^{4C}), 6.48 (s (broad), 2H, N$H_2{}^{2G}$), 6.19 (s, 1H, $H^{1'G}$), 5.77 (s, 1H, $H^{1'C}$), 5.31 (d, J = 6.2 Hz, 1H, $H^{2'G}$), 5.16 (t, J = 5.1 Hz, 1H, $H^{2'C}$), 4.91 (d, J = 6.3 Hz, 1H, $H^{3'G}$), 4.75 (t, J = 5.1 Hz, 1H, $H^{3'C}$), 4.57 (dd, J = 11.3, J' = 6.1 Hz, 1H, $H^{4'G}$), 4.29 (dt, J = 12.9, J' = 6.3 Hz, 5H, H^{5G}, $H^{5'C}$), 4.04 (dd, J = 11.4, J' = 6.9 Hz, 1H, $H^{4'C}$), 2.62 – 2.53 (m, 1H, -COCH), 1.83 (s, 12H, -C$H_3{}^{h,e}$), 1.53 (s, 3H, -OC(CH_3)$_2$), 1.49 (s, 3H, -OC(CH_3)$_2$), 1.31 (s, 3H, -OC(CH_3)$_2$), 1.28 (s, 3H, -OC(CH_3)$_2$), 1.09 (d, J = 6.3 Hz, 15H, -C(CH_3)$_3$, -COCH-(CH_3)$_2$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 177.1, 175.8, 164.1, 156.1, 154.1, 153.1, 150.3, 146.6, 141.0, 139.0, 136.0, 135.1, 130.7, 130.4, 129.0, 121.3, 119.0, 117.4, 113.3, 113.0, 94.0, 93.6, 93.3, 88.8, 85.4, 84.7, 84.4, 83.4, 81.5, 80.7, 78.2, 69.8, 64.2, 63.9, 38.1, 33.1, 27.0, 26.9, 26.8, 25.3, 25.1, 19.1, 18.7, 18.7.

HRMS (MALDI): Calculated for $C_{54}H_{62}N_8O_{12}$: 1015.4565 $[M+H]^+$. Found: 1037.4370 $[M+Na]^+$. Matrix: DCTB

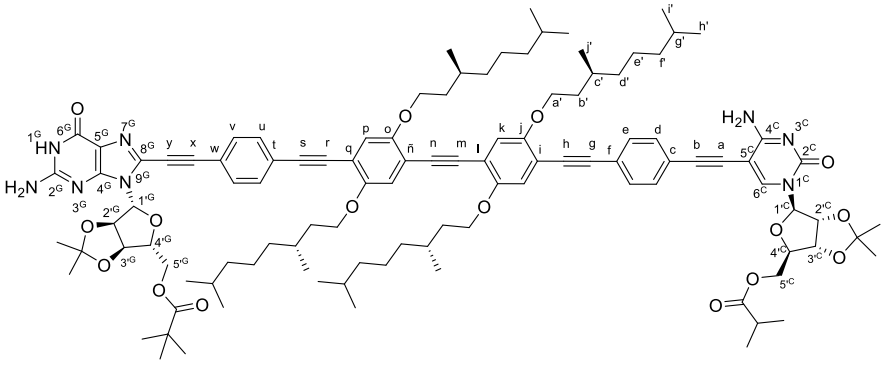


G1-B3-C1. G1-B3-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between **C1-B3-I** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (15 mL) was poured over **C1-B3-I** (190 mg, 0.18 mmol), $PdCl_2(PPh_3)_2$ (2.57 mg, 3.66 µmol), and CuI (0.35 mg, 1.83 µmol). Then, **G1** (94.9 mg, 0.22 mmol), was added dropwise and the mixture was stirred under argon for 12h at 40 ^{o}C. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was reprecipitated in methanol, affording **G1-B3-C1** as a yellow solid (215 mg, 89 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.97 (s (broad), 1H, NH^{1G}), 8.11 (s, 1H, H^{6C}), 7.97 (s (broad), 1H, NH^{4C}), 7.83 – 7.47 (m, 10H, $H^{d,e,o,p}$), 7.24 (s (broad), 1H, NH^{4C}), 7.20 (d, J = 5.4 Hz, 2H, H^{j}), 6.77 (s (broad), 2H, NH_2^{2G}), 6.13 (d, J = 1.4 Hz, 1H, $H^{1'G}$), 5.83 (d, J = 1.8 Hz, 1H, $H^{1'C}$), 5.45 (d, J = 6.3 Hz, 1H, $H^{2'G}$), 5.28 (dd, J = 6.3, J' = 3.3 Hz, 1H, $H^{2'C}$), 5.02 (dd, J = 6.5, J' = 1.8 Hz, 1H, $H^{3'G}$), 4.82 (d, J = 5.6 Hz, 1H, $H^{3'C}$), 4.36 – 4.10 (m, 6H, $H^{4'G}$, $H^{4'C}$, $H^{5'G}$, $H^{5'C}$), 4.06 (t, J = 6.2 Hz, 4H, -(OCH_2)-), 2.64 – 2.55 (m, 1H, -COCH), 1.75 (p, J = 6.7 Hz, 4H, -OCH_2CH_2-), 1.54 (s, 3H, -OC(CH_3)), 1.49 (s, 3H, -OC(CH_3)), 1.34 (s, 3H, -OC(CH_3)), 1.30 (s, 3H, -OC(CH_3)), 1.28 – 1.17 (m, 26H), 1.16 – 1.04 (m, 15H, -COCH-(CH_3)$_2$, -C(CH_3)$_3$), 0.82 (t, J = 6.4 Hz, 6H, (-CH_2CH_3)$_2$.

^{13}C NMR (39 MHz, DMSO-d_6) δ(ppm) = 179.6, 177.4, 164.6, 159.2, 155.0, 154.0, 153.7, 153.5, 150.2, 131.7, 131.4, 131.1, 131.0, 124.2, 123.7, 120.6, 120.0, 118.5, 116.8, 114.6, 114.0, 114.0, 113.7, 96.1, 95.1, 94.2, 92.8, 90.2, 88.8, 88.3, 87.2, 85.3, 85.1, 84.7, 83.2, 81.3, 79.6, 79.1, 69.6, 69.3, 66.4, 64.8, 53.4, 38.9, 33.9, 32.0, 31.9, 29.7, 29.5, 29.3, 27.3, 27.2, 27.1, 26.2, 25.4, 25.2, 22.8, 22.7, 19.0, 14.2.

HRMS (ESI+): Calculated for $C_{76}H_{90}N_8O_{14}$: 1339.6649 $[M+H]^+$. Found: 1339.6638 $[M+H]^+$**.**



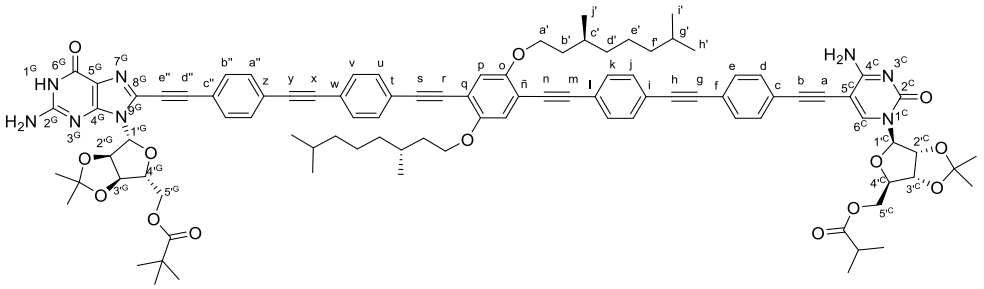
G1-B4-C1. G1-B4-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between **C1-B4-I** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (8 mL) was poured over **C1-B4-I** (45 mg, 29 µmol), $PdCl_2(PPh_3)_2$ (0.42 mg, 0.6 µmol), and CuI (0.06 mg, 0.3 µmol). Then, **G1** (26 mg, 0.06 mmol), was added dropwise and the mixture was stirred under argon for 12h at 40 ^{o}C. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1), affording **G1-B4-C1** as a yellow solid (46 mg, 85 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.97 (s (broad), 1H, NH^{1G}), 8.11 (s (broad), 1H, H^{6C}), 7.98 (s (broad), 1H, NH^{4C}), 7.82 – 7.44 (m, 8H, $H^{d,e,v,u}$), 7.22 (s (broad), 1H, NH^{4C}), 7.19 (d, J = 5.9 Hz, 2H, $H^{p,j}$), 7.06 (s, 2H, $H^{o,k}$), 6.78 (s (broad), 2H, NH_2^{2G}), 6.13 (s, 1H, $H^{1'G}$), 5.83 (s, 1H, $H^{1'C}$), 5.44 (d, J = 6.3 Hz, 1H, $H^{2'G}$),

5.29 (d, *J* = 4.2 Hz, 1H, $H^{2'C}$), 5.01 (d, *J* = 6.3 Hz, 1H, $H^{3'G}$), 4.91 – 4.74 (m, 1H, $H^{3'C}$), 4.17 (dq, *J* = 61.7, *J'* = 6.9 Hz, 14H, $H^{4'G}$, $H^{4'C}$, $H^{5'G}$, $H^{5'C}$, -(OCH_2)-), 2.61 (dd, *J* = 13.7, *J'* = 6.8 Hz, 1H, -COC*H*), 1.90 – 1.01 (m, 67H, -C$H_2$$^{b'-g'}$-, OC(C$H_3$), -COCH-(C$H_3$)$_2$,-C(CH_3)$_3$), 0.93 (d, *J* = 6.1 Hz, 12H, -C$H_3$$^{j'}$), 0.78 (dd, *J* = 6.6, *J'* = 4.0 Hz, 24H, -C$H_3$$^{i',h'}$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 177.0, 175.6, 164.0, 156.1, 154.1, 153.1, 153.0, 152.9, 152.9, 150.2, 131.7, 131.4, 131.3, 130.9, 128.7, 124.0, 122.5, 120.2, 117.5, 116.7, 116.3, 114.0, 113.8, 113.2, 113.0, 112.9, 112.8, 94.2, 93.9, 93.7, 93.6, 92.6, 91.6, 91.5, 88.9, 88.1, 85.5, 84.7, 84.4, 83.5, 81.5, 80.7, 80.4, 67.4, 67.1, 64.2, 63.8, 38.1, 36.6, 36.5, 35.7, 35.7, 33.1, 29.3, 29.3, 27.3, 26.9, 26.9, 26.7, 25.2, 25.0, 24.2, 24.1, 22.4, 22.3, 19.5, 19.4, 19.3, 18.7, 18.6.

HRMS (ESI+): Calculated for $C_{108}H_{142}N_8O_{16}$: 1807.0538 $[M]^+$. Found: 1807.0458 $[M]^+$.



G1-B5-C1. G1-B5-C1 was prepared according to a *Standard Procedure C* for the Sonogashira coupling reaction between **C1-B5-I** and the ethynyl-nucleobase **G1**. A dry a THF/NEt_3 (4:1) mixture (6 mL) was poured over **C1-B5-I** (35 mg, 27 µmol), $PdCl_2(PPh_3)_2$ (0.38 mg, 0.54 µmol), and CuI (0.05 mg, 0.3 µmol). Then, **G1** (23 mg, 0.05 µmol), was added dropwise and the mixture was stirred under argon for 12h at 40 °C. Once completed, the mixture was filtrated over a celite plug and the solvent was evaporated under reduced pressure. The product was was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1), affording **G1-B5-C1** as a yellow solid (26 mg, 61 %).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.99 (s (broad), 1H, NH^{1G}), 8.12 (s, 1H, H^{6C}), 7.99 (s (broad), 1H, NH^{4C}), 7.84 – 7.40 (m, 16H, $H^{d,e,j,k,u,v,a'',b''}$), 7.25 (s, 1H, NH^{4C}), 7.19 (s, 2H, $H^{p,o}$), 6.78 (s (broad), 2H, NH_2^{2G}), 6.13 (s, 1H, $H^{1'G}$), 5.82 (s, 1H, $H^{1'C}$), 5.44 (d, *J* = 6.3 Hz, 1H, $H^{2'G}$), 5.34 – 5.23 (m, 1H, $H^{2'C}$), 5.02 (d, *J* = 6.3 Hz, 1H, $H^{3'G}$), 4.90 – 4.75 (m, 1H, $H^{3'C}$), 4.47 – 3.94 (m, 10H, $H^{4'G}$, $H^{4'C}$, $H^{5'G}$, $H^{5'C}$, -(OCH_2)-), 2.66 – 2.54 (m, 1H, -COC*H*), 1.87 – 1.01 (m, 47H, -C$H_2$$^{b'-g'}$-, OC(C$H_3$), -COCH-(C$H_3$)$_2$,-C(CH_3)$_3$), 0.93 (d, *J* = 6.1 Hz, 6H -C$H_3$$^{j'}$), 0.79 (d, *J* = 6.5 Hz, 12H, -C$H_3$$^{i',h'}$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 177.1, 175.8, 164.0, 156.1, 154.2, 153.1, 150.3, 147.1, 132.7, 132.4, 131.9, 131.8, 131.7, 131.6, 131.4, 131.4, 128.7, 123.3, 123.1, 122.9, 122.8, 122.1, 121.9, 121.8, 121.5, 120.5, 117.6, 116.5, 113.3, 113.1, 113.0, 94.3, 93.6, 92.6, 91.5, 91.0, 90.8, 89.8, 88.8, 88.5, 85.5, 84.7, 84.4, 83.4, 81.5, 80.7, 80.6, 67.2, 64.2, 63.9, 38.1, 36.6, 35.7, 33.1, 29.3, 27.3, 27.0, 26.9, 26.7, 25.3, 25.1, 24.2, 22.4, 22.3, 19.4, 18.7.

HRMS (ESI+): Calculated for $C_{96}H_{106}N_8O_{14}$: 1595.7907 $[M+H]^+$. Found: 1617.7716 $[M+Na]^+$.

5.5.2 NMR and Optical Spectroscopy Dilution and Titration Experiments.

The NMR and optical spectroscopy dilution and titration experiments detailed in the Experimental Section **1.4** of *Chapter 1* and Experimental Section **2.3** of *Chapter 2* and Experimental Section **3.3.2** of *Chapter 3* are also applicable here

Summary & Conclusions.

Introduction.

The impressive field of ***Supramolecular Chemistry***, developed in the last decades, has served as a powerful and versatile tool for the high-fidelity construction of materials and systems assembled through the sum of weak ***non-covalent interactions***, which can become much stronger *via* cooperative and multivalent effects. Supramolecular chemistry is thus considered as a great option to construct, *via* a bottom-up approach, target nanostructures from precisely designed molecules that carry the necessary information (size, shape and functionality) to self-assemble spontaneously.

A molecule with more than one binding site may assemble into linear (open) or cyclic (closed) structures in process called ***Ring-Chain Equilibria***. Although the size of linear oligomers can be sometimes limited within a certain range, the supramolecular product is commonly a statistical distribution of chain lengths. Therefore, the synthesis of well-defined, discrete supramolecular structures has normally been focused on closed (multi)macrocyclic systems, where size is dictated by the geometric requirements of both the monomer and the binding interaction or by the use of particular templates, both of these approaches leading to ring closure. The resulting cyclic species may be formed quantitatively because it enjoys a thermodynamic stability that is substantially larger than the sum of the corresponding individual interactions. The effect that causes such increased stability is defined as ***Chelate Cooperativity*** and stems from the fact that an intramolecular interaction is favoured over an intermolecular one, providing a series of conditions of enthalpic and entropic origin are met. The increased in stability when comparing a linear and a cyclic oligomer of a certain length is given by the product K_{inter}·*EM*, where K_{inter} is the intermolecular binding constant and considers the additional association to form the cycle, whereas *EM*, a key parameter quantifying chelate cooperativity, stands for ***Effective Molarity*** and takes into account that the last binding event to form the cycle is intramolecular ($EM = K_{intra}/K_{inter}$). It is now well recognized that rigid monomers, having a preorganized structure that affords unstrained complexes, are most suited to produce high *EM*s and thus quantitative yields upon self-assembly in discrete nanoobjects.

General Objective.

In this Thesis, we have investigated the formation of well-defined cyclic systems in solution by H-bonding assembly of chemically programmed monomers. By the rational use of the tools of molecular self-assembly, cooperative phenomena, and from the knowledge acquired on the supramolecular properties provided by DNA nucleobase derivatives, we have achieved a rigorous control and fidelity at the nanoscale. In this way, we have used the high selectivity and directionality of triple H-bonding Watson-Crick interactions for constructing discrete, stable cyclic tetramers in solution from four ditopic monomeric subunits that are able to self-assemble even in polar media (Figure 1).

The monomers basically comprise a *rigid π-conjugated* ***central block*** that is linearly disubstituted at both ends through suitable ***spacers*** with *self-assembling* ***directors,*** able to interact by H-bonding.

The H-bonding ***directors*** are actually nucleobases, *i.e.* naturally occurring guanine (G), cytosine (C) and uracil (U); and non-natural 2-aminoadenine (A), *iso*guanine (iG) and *iso*cytosine (iC). Together, they constitute a family of three complementary Watson-Crick couples: G–C, A–U, and iG–iC, that associate *via* triple H-bonding patterns. When these bases

are subjected to electrophilic halogenation reactions, purines are substituted at the 8-position and pyrimidines at the 5-position (marked as **X** in Figure 1). It is essential to note that when the purine-pyrimidine pairs interact *via* complementary Watson-Crick H-bonding, these two positions form an exact 90º angle. On the other hand, the marked **R^1** positions, the 9-position in purines and the 1-position in pyrimidines, are always pointing toward the exterior of the nucleobase pairs.

The ***central blocks*** have to be rigid units substituted at both ends with an exact angle of 180º. They will typically comprise π-conjugated units that carry reactive groups at opposite positions and that can be endowed with diverse functions and equipped with different lateral groups. There is a large number of functional π-conjugated units that fulfil the prerequisite of linear di-substitution: TTF, conjugated oligomers, aromatic acenes, perylenes, ABAB porphyrin or phthalocyanine macrocycles, etc.

Ethynyl groups have been selected as the ***spacers*** between the different units forming the monomer. Such groups are interesting to us because they are linear, rigid and allow for substitution in this suitable 180º angle. They have some rotational liberty around the connecting σ-bonds and a minimum steric hindrance, which is necessary for sufficient conformational freedom between bases and central blocks. They are also π-conjugated and can facilitate electronic coupling between the different units in the cyclic tetramers.

Overall, the linear structure of the dinucleobase monomer together with the 90º angle imposed by Watson-Crick H-bonding interactions at the edges will lead to the formation of cyclic, rectangular assemblies composed of four molecules.

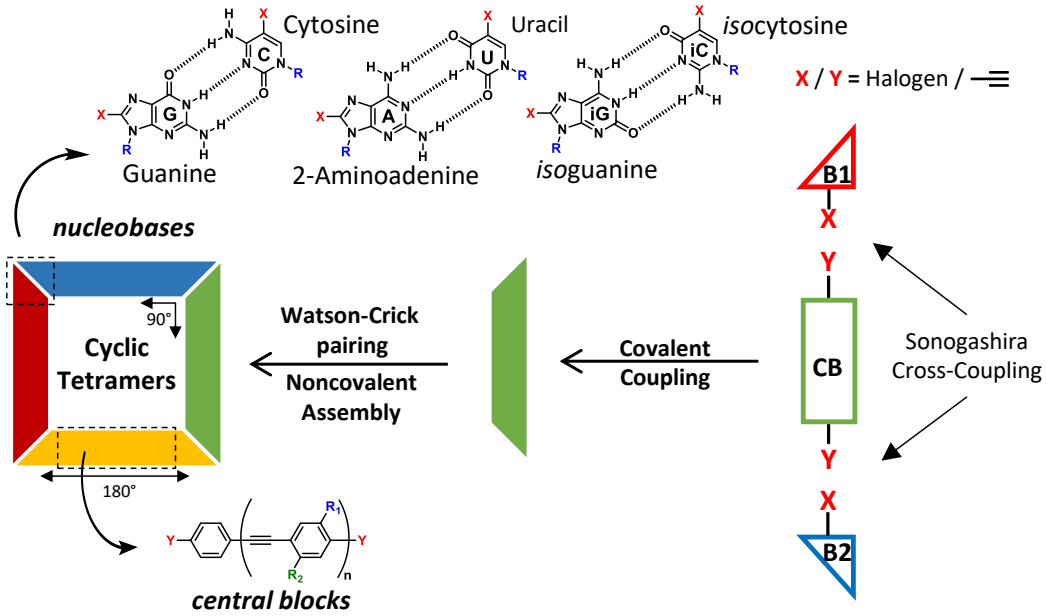


Figure 1. Molecular components and self-assembly strategy toward cyclic tetramers. Central Block (**CB**), Nucleobase (**B**).

Chapter 1. Monomer Design and Synthesis.

The target monomers share a common structure based on rigid and planar π-conjugated central blocks that are linearly disubstituted at both ends with nucleobase directors. In *Chapter 1* we have firstly focused on designing and synthesizing the individual molecular components. In particular, we have developed new optimized strategies to synthesize a wide family of dihalogenated central blocks as the core of the monomer, as well as to prepare a complete variety of nucleobase derivatives substituted with an ethynyl moiety. A proper synthetic strategy has also been drawn up in this Chapter in order to reach the final ditopic monomers by conveniently coupling the building blocks in high yields through palladium-catalyzed Sonogashira reactions. In this first purely synthetic task, a wide family of both the mentioned nucleobase derivatives and dihalogenated central blocks has been synthesized as precursors of the final target molecules that will be used for further studies in the next chapters (Figure 2).

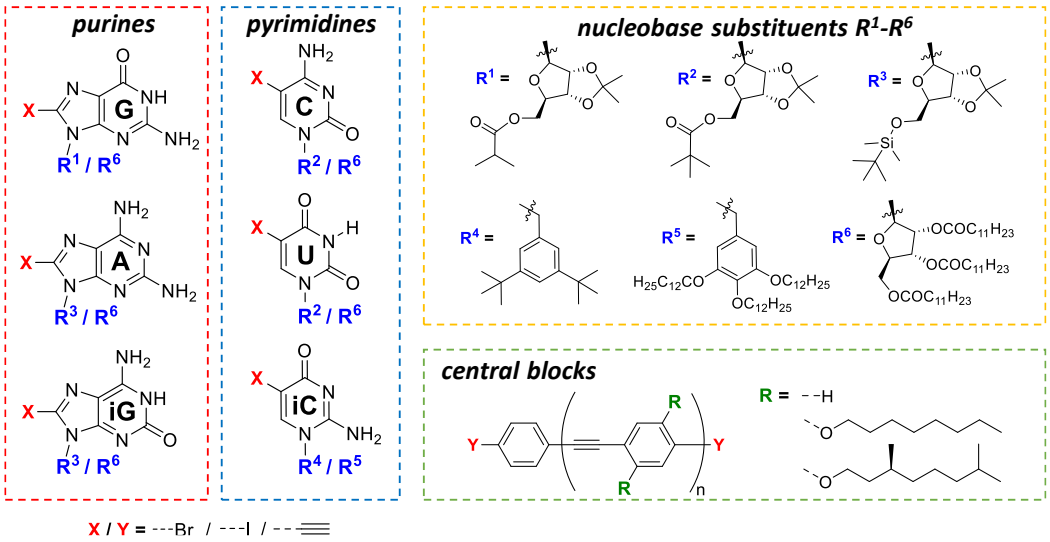


Figure 2. Molecular components considered and synthesized in *Chapter 1*.

Five different π-conjugated molecules formed by oligophenylene-ethynylene cores with alkyl chains have been synthesized to complete a useful collection of seven dihalogenated central blocks to be used for different purposes in this Thesis, as has been detailed in each of the subsequent Chapters. These building blocks have been designed with the aim to afford the necessary rigidity (π-conjugated body) and solubility (lateral alkyl chains) to the target dinucleoside monomers that can associate in cyclic species with a variable size.

On the other hand, a series of lipophilic nucleobases have been prepared, comprising natural and non-natural derivatives, which are substituted at the 5-(pyrimidines) or 8-position (purines) with either a halogen atom or a terminal triple bond. These include cytidine (**C**), *iso*cytosine (**iC**) and uridine (**U**) as pyrimidine derivatives and guanosine (**G**), *iso*guanosine (**iG**)and 2-aminoadenosine (**A**) as a complementary purine bases. The ribose moiety has been equipped with different bulky groups to afford solubility in the most commonly employed organic solvents and prevent π-π interactions. In another set of derivatives, the ribose groups have been functionalized with long alkyl chains in order to improve further the solubility in more apolar organic solvents, such as toluene or carbon tetrachloride. In other products, like

iC derivatives, the base was instead substituted with two different benzyl groups (R^4 and R^5 in Figure 2). The synthetic pathway leading to the final ethynylated compounds have been optimized for each nucleobase attending to reactivity problems, convenience, ease of purification, and overall yields. The final results indicate that the choice of the synthetic route is not trivial, and that each nucleobase requires a particular optimized protocol to reach the final halogenated/ethynylated products.

Chapter 2. Evaluation of Dimerization and Association Constants between Lipophilic Mononucleotides.

In *Chapter 2* we have prepared a new series of lipophilic nucleosides comprising natural and non-natural bases that are π-conjugated to a short oligophenylene-ethynylene fragment. These bases include guanosine, *iso*guanosine, 2-aminoadenosine as purine heterocycles, and cytidine, *iso*cytosine and uridine as complementary pyrimidine bases (Figure 3). The H-bonding dimerization and association processes between complementary bases were evaluated using different techniques (^{1}H NMR and absorption spectroscopies), solvents (toluene, $CHCl_3$:CCl_4 (2:3), $CHCl_3$, THF and DMF), concentration ranges, and fitting programs.

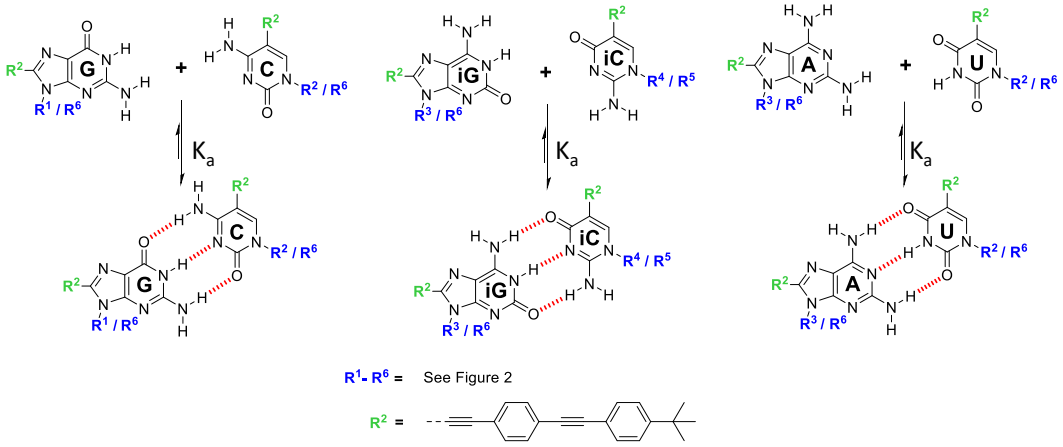


Figure 3. Molecular components considered and synthesized in *Chapter 2*.

Considering $CHCl_3$ (or $CDCl_3$) as a standard solvent to compare binding strength, symmetric *ADA-DAD* H-bonding patterns (**A**-**U** base pairs) afford 1:1 binding constants in the order of 10^2 M^{-1}, whereas unsymmetric *DDA-AAD* H-bonding patterns (**G**-**C**, and the novel **G**-**iC**, **iG**-**C** and **iG**-**iC** base pairs) yield association constants in the order of 10^4 M^{-1}. Such an increase, in approximately two orders of magnitude, is well-known in the literature and is caused by the establishment of stabilizing secondary H-bonding interactions in the *DDA-AAD* pairs, as explained in the Introduction of this Thesis. Binding constants between mononucleosides are however very sensitive to the polarity of the solvent, and they increase in about one order of magnitude in more apolar solvents, like toluene or $CHCl_3$:CCl_4 mixtures, and decrease when the solvent can compete strongly for the H-bonding sites, like in THF or DMF.

This work provides, to the best of our knowledge, the first association constant values between complementary nucleobases in diverse solvents like toluene, THF or DMF. It also includes K_a calculations for the non-natural **iG**-**iC** pair and all possible combinations of the

purine-pyrimidine pairs between **G**, **C**, **iG** and **iC** bases. All of these combinations reveal comparable K_a values as a consequence of the similar *DDA-AAD* triple H-bonding patterns.

Most of the work presented in *Chapters 1 and 2* was published in the following article: J. Camacho-García, C. Montoro-García, A. M. López-Pérez, N. Bilbao, S. Romero-Pérez, D. González-Rodríguez. "Synthesis and Complementary Self-association of Novel Lipophilic π-Conjugated Nucleoside Oligomers". *Org. Biomol Chem.* **2015**, *13,* 4506-4513 (10.1039/C5OB00098J).

Chapter 3. Cyclic Tetramer Self-assembly in Solution.

In *Chapter 3* we focused on the study of the H-bonding cyclotetramerization process in solution of ditopic monomers, bearing self-complementary nucleobases (G-C, iG-iC and A-U) at their edges. We investigated the fidelity of this self-assembly process, where association into discrete cyclic tetramers, instead of open oligomeric structures or other kind of strained structures, should be highly favoured (Figure 4).

The design of these monomers thus relies on the optimization of the *chelate cooperativity,* and thus on obtaining high *effective molarity* values in their cyclization process. For this purpose, a series of ditopic monomers, carrying complementary nucleobase derivatives on each side of a linear and rigid *p*-phenylene-ethynylene group (**G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-C1**), has been prepared and their self-assembly in diverse solvents studied through concentration- and temperature-dependent measurements, as well as through denaturation experiments. It should be underlined that, aside from the exceptional thermodynamic stability shown by the *c***G1-B1-C1**$_4$ / *c***iG1-B1-iC1**$_4$ / *c***A1-B1-U1**$_4$ macrocycles studied in *Chapter 3*, which can resist high dilutions and polar environments, they constitute a kinetically stabilized product in the overall self-assembly landscape. No matter the concentration, temperature or solvent conditions employed, the cyclic tetramers are always observed as slowly exchanging species in the NMR timescale (Figure 5), exhibiting exchange rate constants within the 1-10 s^{-1} range.

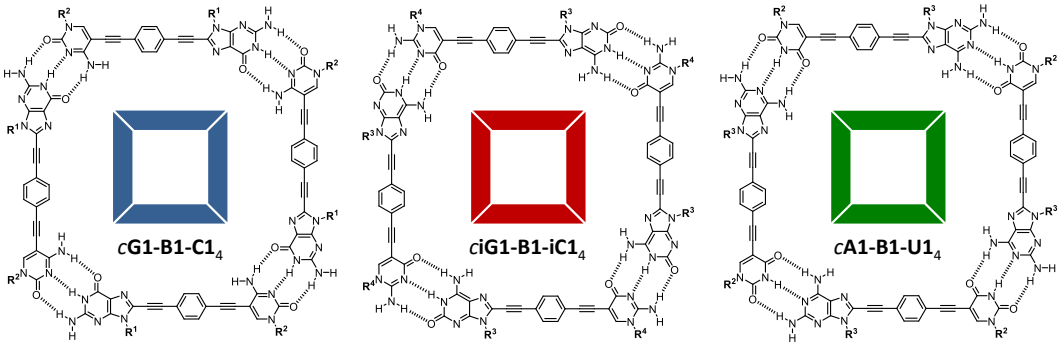


Figure 4. Structure of lipophilic dinucleoside **G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-C1** considered in *Chapter 3*.

Furthermore, we have examined in this Chapter the role of the symmetry of multipoint H-bonding on chelate cooperativity in supramolecular macrocyclization processes. We have been able to dissect and analyze independently the contributions of the H-bonding strength between complementary nucleobases, explained by the Jorgensen model, and the intrinsic chelate effect that they exert in cyclic systems. The results presented clearly demonstrate that cyclic systems constructed from symmetric *DAD-ADA* H-bonding pairs are much less stable than the homologues assembled from unsymmetric *ADD-DAA* or *DDA-AAD* pairs. On one hand, the *DAD-ADA* bonding pattern reduces considerably the enthalpy of intermolecular association due to the absence of attractive secondary interactions between vicinal H-bonding groups. On the other, the symmetry of this pattern introduces the possibility of multiple binding modes and hence a higher number of degrees of freedom in the competing linear oligomers, which must be lost upon cyclization. This effect, of entropic origin, has a large impact on the *EM* of the system, which in our case is reduced by about 3 orders of magnitude (from *EM* = 10^2-10^3 M for *c***G1-B1-C1**$_4$ / *c***iG1-B1-iC1**$_4$ to *EM* = 0.1-1 M for *c***A1-B1-U1**$_4$).

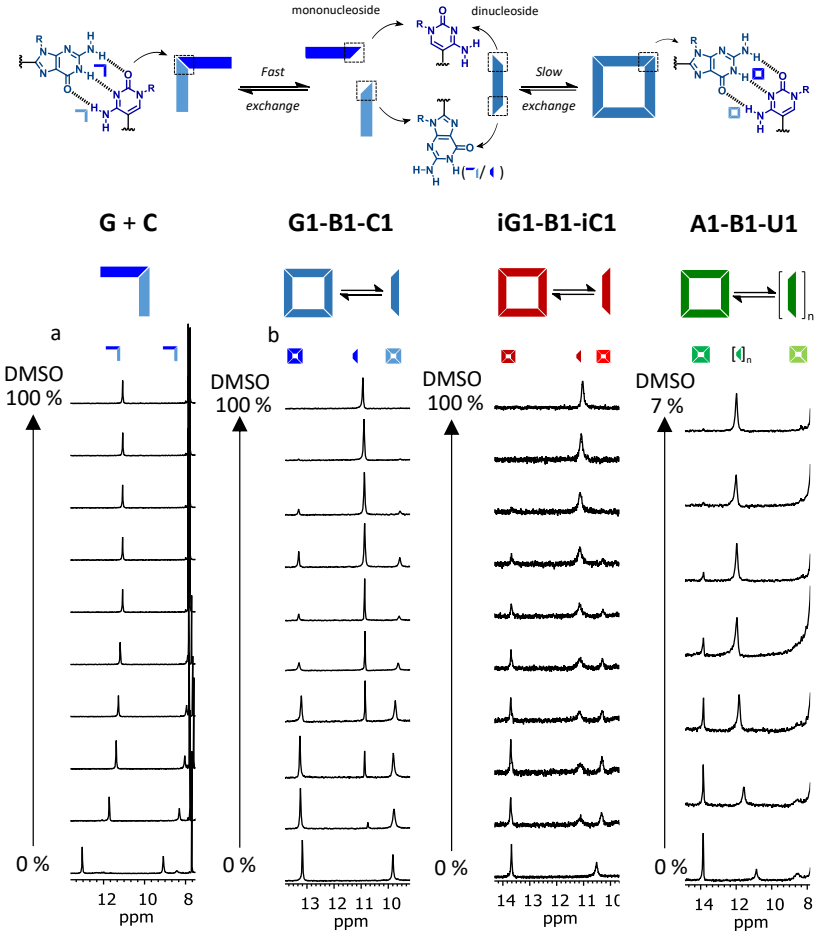


Figure 5. *Denaturation experiments with increasing amounts of DMSO.* 14-8 ppm region of the ^{1}H NMR spectra of (**a**) **G**+**C** 1:1 complex and (**b**) **G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-U1** upon increasing the volume fraction of DMSO-d_6 in $CDCl_3$. (C = 1 x 10^{-2} M, T = 298 K in all cases). On the top of the image we can observe two different scenarios of the self-assembly phenomena between mononucleosides and dinucleosides, which exhibit thermodynamic and kinetic differences. On one hand, **G1**+**C1** H-probes in $CDCl_3$ confirm the association between nucleobases, whilst increasing amounts of DMSO-d_6 content that competes highly for

H-bonding, provokes an upfield **G1** amide signal that is attributed to a fast-exchange equilibria in the ^{1}H NMR timescale between dimers and monomers due to nucleobases are lees involved in intermolecular H-bonding interaction. On the other hand, c**G1-B1-C1**$_4$ shows a high resistance to this highly polar cosolvent like c**iG1-B1-iC1**$_4$, and the cyclic tetramer persisted even after the addition of ca. 80% DMSO-d_6. The c**A1-B1-U1**$_4$ assembly, on the contrary, could not resist more than 7% DMSO-d_6 in $CDCl_3$. Also these measurements revealed the presence of an equilibrium between monomer and cyclic tetramer. In all cases, upon addition of polar solvent, monomer grows at the expense of tetrameric species. It is interesting to note that the shape and position of the **G** (or **U**)-amide and **C** (or **A**)-amine do not change with concentration, suggesting a very low exchange in the NMR timescale and an "all or nothing" behavior where the tetrameric species are by far the most stable cyclic assemblies in solution or nothing else can survive and the monomeric species in the only entity present.

Our conclusions could in principle be extended to many linear or cyclic supramolecular systems assembled *via* multipoint binding interactions. If a discrete, well-defined closed architecture is to be designed, rigid monomers with a suitable geometry in combination with an unsymmetric binding motif should be used to enhance the *EM*s of the cyclic system(s). In other words, the binding interaction should also contribute to the preorganization of the system towards a specific cycle, reducing the degrees of freedom of any other competitive supramolecular species. If, on the other hand, linear supramolecular polymers are pursued, a symmetric multipoint binding interaction would be the best choice to minimize chelate cooperativity, and hence the tendency of the supramolecular system to form undesired cycles.

Most of the work presented in *Chapter 3* was published in the following articles:

C. Montoro-García, J. Camacho-García, A. M. López-Pérez, N. Bilbao, S. Romero-Pérez, M. J. Mayoral, D. González-Rodríguez. "High-fidelity Noncovalent Synthesis of Hydrogen-bonded Macrocyclic Assemblies". *Angew. Chem. Int. Ed.* **2015***, 54,* 6780-6784 (10.1002/anie.201501321). (VIP article).

C. Montoro-García, J. Camacho-García, A. M. López-Pérez, M. J. Mayoral, N. Bilbao, D. González-Rodríguez. "Role of the Simmetry of Multipoint Hydrogen Bonding on Chelate Cooperativity in Supramolecular Macrocyclization Processes". *Angew. Chem. Int. Ed.* **2016**, *55,* 223-227 (10.1002/anie.201508854).

Chapter 4. Self-sorting Phenomena Governed by Chelate Cooperativity.

Self-sorting phenomena has been studied in *Chapter 4*, in which narcissistic macrocyclization behaviour of our three dinucleoside monomers has been inquired by comparing the self-assembly in solution of their binary mixtures with quaternary mixtures of the corresponding mononucleosides (Figure 6a).

Because of the different, self-complementary *DDA-AAD* / *DAD-ADA* Watson-Crick H-bonding patterns of the **G1-B1-C1** + **A1-B1-U1** and the **iG1-B1-iC1** + **A1-B1-U1** combinations, we initially expected them to self-sort, and thus self-assemble separately in their respective cyclic tetramers. But in the case of the **G1-B1-C1** + **iG1-B1-iC1** combination, an unsymmetric *DDA-AAD* H-bond pattern is shared and the observation of self-sorting phenomena is not so clear. Since, as established in *Chapter 2*, G may bind to iC and iG to C through reverse Watson-Crick pairs with comparable H-bonding strength than the regular G-C and iG-iC Watson-Crick pairs, the association of these two dinucleoside molecules might lead instead to a complex mixture of linear and cyclic oligomers, reducing in this way the fidelity of the cyclotetramerization process.

As a matter of fact, as it is depicted in Figure 6, when monomers **G1-B1-C1** and **A1-B1-U1** are mixed, the characteristic signals of their respective one-component tetramers appeared separately, which is indicative of self-sorting. However, in a similar way, when **G1-B1-C1** and **iG1-B1-iC1** monomers are mixed, the signals of their corresponding macrocycles are again found separately in the ^{1}H NMR spectra, and no sign of any other associated or non-associated species was detected. On the other hand, NOESY experiments of both mixtures further confirm that A only binds to U, G only binds to C, and iG only binds to iC in these two dinucleoside combinations.

These results clearly confirm the narcissistic self-sorting behaviour between **G1-B1-C1** and **A1-B1-U1** monomers, which was an expected behaviour, and between **G1-B1-C1** and **iG1-B1-iC1**. However, in the last case, self-sorting phenomena is not governed by the H-bonding pattern, but rather by the high *chelate cooperativity* exhibited by their respective cyclic tetramers, as determined in *Chapter 3*, which predicts that the most stable self-assembled structure will be the smallest and the least strained cycle. Cyclic tetramers can only be produced when the regular G-C and iG-iC Watson-Crick pairs, displaying a 90° angle between the 8-purine and 5-pyrimidine positions, are established. This is a new driving force that directs the self-sorting processes in supramolecular chemistry that will be taken into account to develop more sophisticated mixture assemblies.

Most of the work presented in *Chapter 4* will be published in the following article, which is under preparation:

“Self-sorting Phenomena Governed by Chelate Cooperativity”, C. Montoro-García, D. Serrano-Molina, M. J. Mayoral, D. González-Rodríguez.

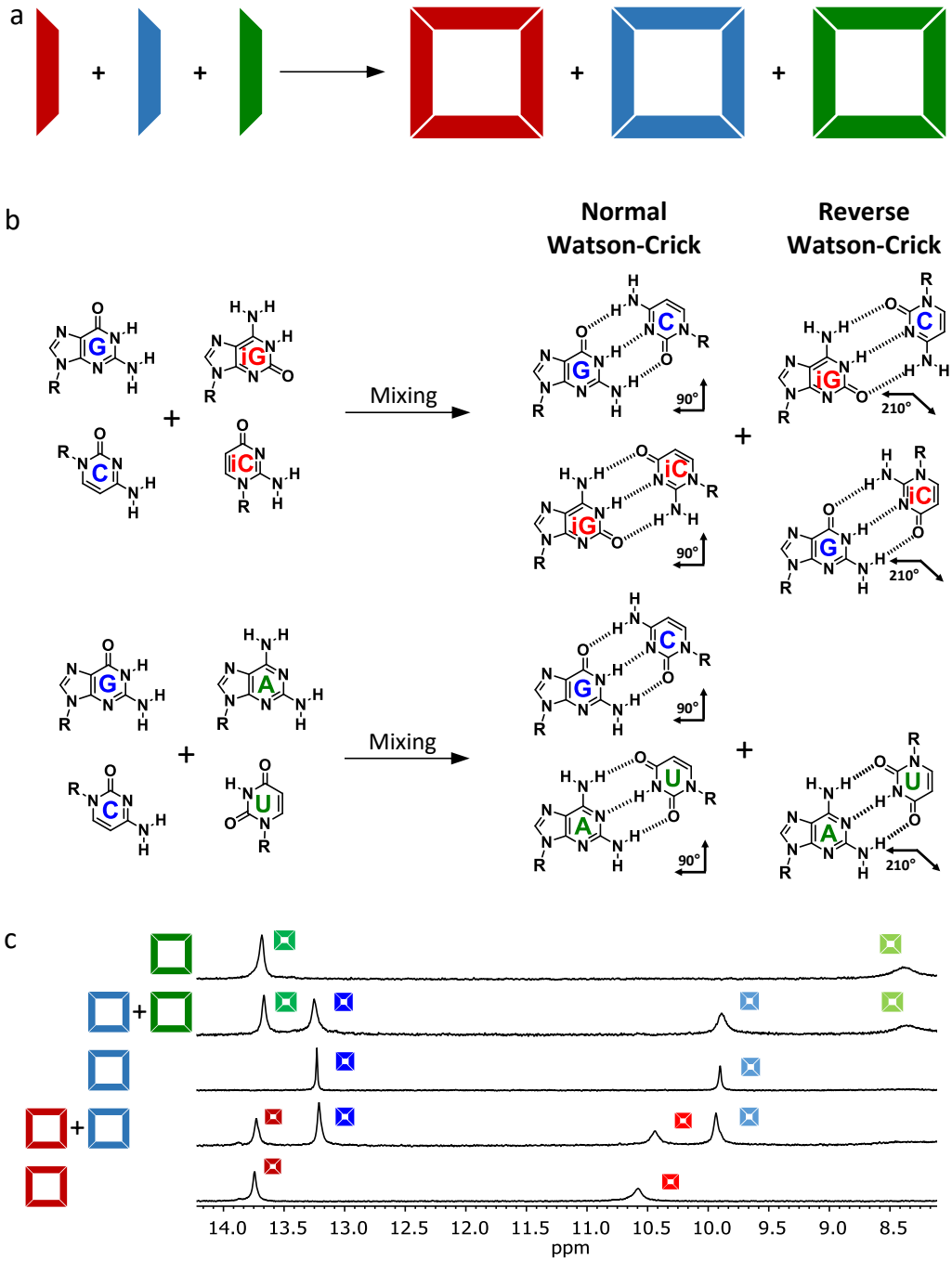


Figure 6. *Self-sorting phenomena of the ditopic monomers.* (**a**) Schematic representation of Self-Sorting phenomena, (**b**) different Watson-Crick H-bonded base-pairs studied in this Chapter and (**b**) downfield region of the ^{1}H NMR spectra showing the H-bonded G-amide (H^{1}; dark blue) and C-amine (H^{2}; light blue), U-imide (H^{1}; dark green) and A-amine (H^{2}; light green), iG-amide (H^{1}; dark red) and iC-amine (H^{2}; light red) proton signals of **G1-B1-C1, A1-B1-U1**, and **iG1-B1-iC1** and their mixtures.

Chapter 5. Impact of Ring Size on Chelate Cooperativity in Noncovalent Macrocyclizations.

Once the macrocyclization process and self-sorting phenomena of **G1-B1-C1**, **iG1-B1-iC1** and **A1-B1-U1** were studied, *Chapter 5* is devoted to assess which is the impact on the chelate cooperativity when the length of the central block is lengthened or shortened. These modifications of the central block vary from a simple C-C bond to a penta-(phenylene-ethynylene) π-conjugated structure (Figure 7). All of these central blocks are substituted at both edges by the same G-C nucleobase pair because, as we explained previously, this unsymmetric base-pairing motif enhances the chelate effect and is the best way to compare the stability that each central block confers to the tetrameric assembly.

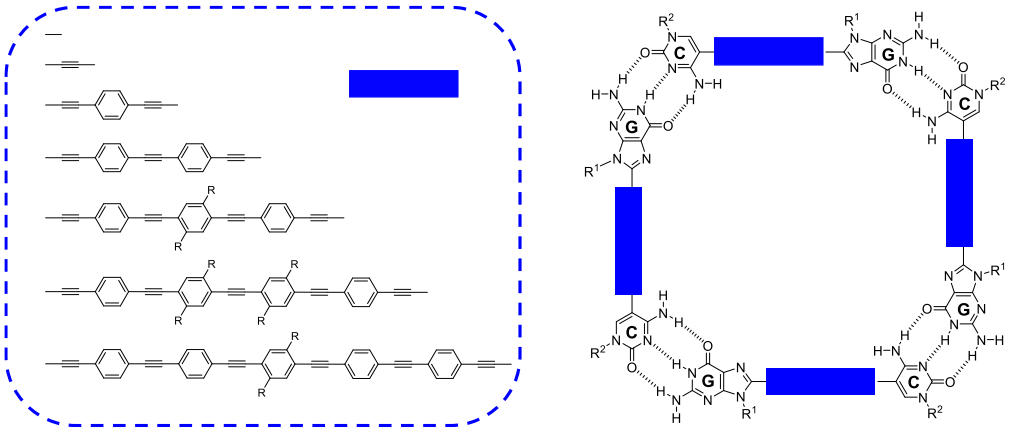


Figure 7. Central blocks with different lengths between complementary G-C base-pair.

The studies and the analyses of the self-assembly of the diverse monomers through concentration- and temperature-dependent measurements, as well as by means of denaturation experiments with either a polar cosolvent (DMSO) or the complementary C mononucleoside, clearly demonstrate the impact that each central block placed between complementary nucleobases exerts on the chelate cooperativity of the cyclic system in two different ways.

Firstly, when the central block becomes larger than the classic *p*-phenylene-diethynylene block studied in *Chapter 3*, with monomer lengths ranging from 2.1 to 4.7 nm, there is a notable loss in the value of the effective molarity that can encompass 5 orders of magnitude. We found that this effect is only due to entropic factors, while the enthalpic component in the cyclotetramerization reaction remains constant for all G-C molecules, since monomer geometry and binding interaction is the same. Our explanation is that, as the number of *σ*-bonds that can more freely rotate, bend and torsion in this central oligophenylene-ethynylene spacer increases, additional degrees of freedom are reached in the monomer and linear oligomers that must be lost upon cyclization (Figure 8). Both ΔS and Ln *EM* seem to follow linear relationships with the number of *σ*-bonds that are independent on the solvent employed. The extrapolation of these trends affords an estimation on how large can we build a cyclic tetramer assembly using this kind of monomers. Any change to the oligo(phenylene-ethynylene) structure of the spacer may lead to important deviations from this general trend, as demonstrated with related monomers equipped with biphenyl spacers.

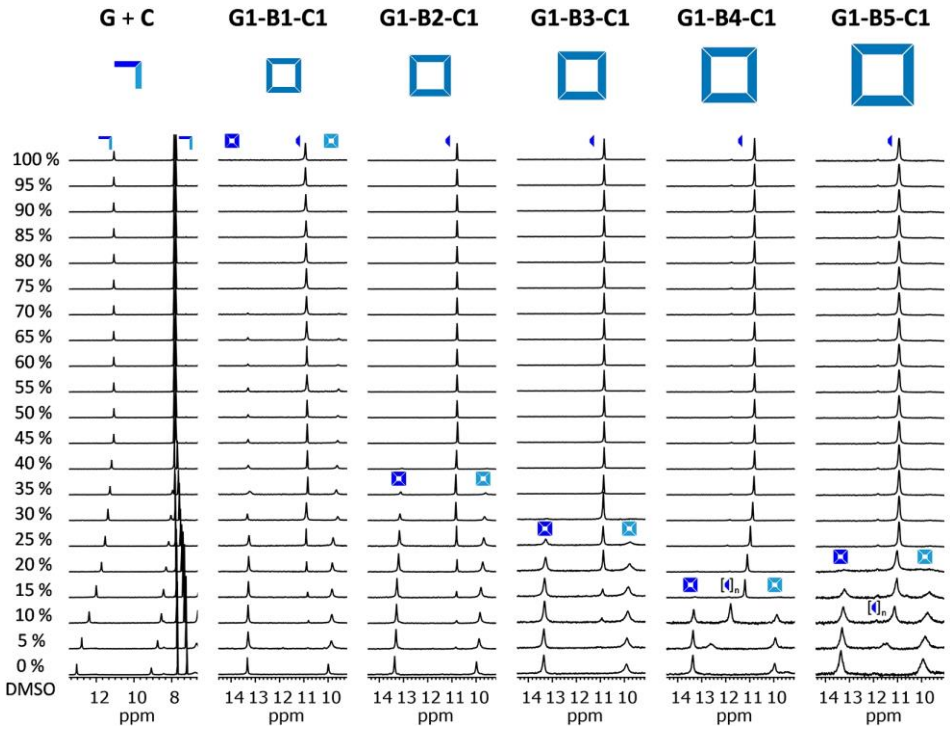


Figure 8. *Denaturation experiments with increasing amounts of DMSO.* Downfield region of the ^{1}H NMR spectra, showing the H-bonded G-amide (H^1; dark blue) and C-amine (H^2; light blue) proton signals, of a 1:1 **G**+**C** mixture and of **G1-B1-C1** – **G1-B5-C1** dinucleosides as the volume fraction of DMSO-d_6 is increased in $CDCl_3$-DMSO-d_6 mixtures at C = 1.0x10^{-2} M and T = 298 K. Additional spectra were recorded at low DMSO content for **G1-B2-C1** – **G1-B5-C1**, but are not shown here for the sake of clarity and homogeneity.

Secondly, when the central linker becomes shorter, as a triple bond or a single C-C bond, other effects become important. Steric hindrance is one of them, and has an important role in order to reach the desired "*syn*" conformation between Watson-Crick edges that is needed to form the cyclic tetramer. The steric volume and the direction in which the substitution of the nucleobases is pointing is crucial (Figure 9) to decide whether these "shorter" molecules will assemble as supramolecular polymers, thus yielding viscous gels in numerous solvents, or discrete closed assemblies.

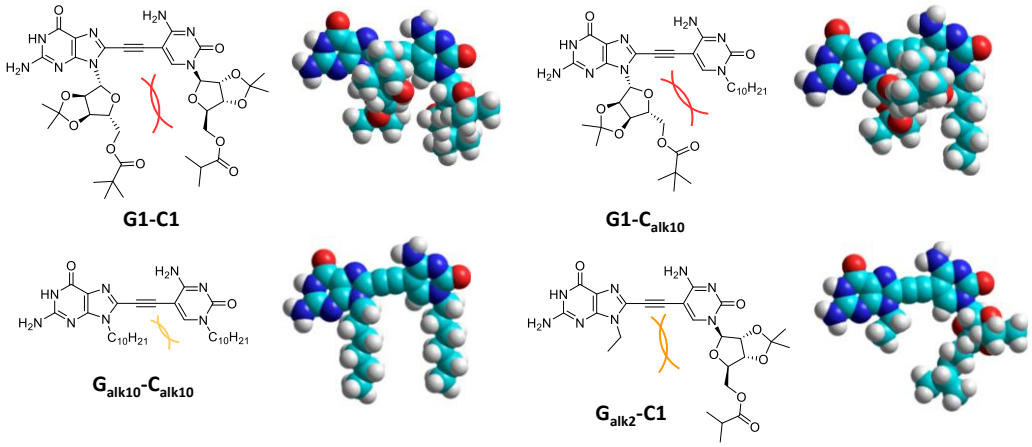


Figure 9. Steric hindrance and molecular model (right) of the different ethynyl-linked monomers synthesized in *Chapter 5*. Colour indicate the degree of the steric hindrance, red (strong) and orange (moderate).

Our conclusions could in principle be extended to many supramolecular cycles or cages in which monomer size is lengthened or shortened. Thus, the previous design of the respective building blocks and the substituents that confer solubility in organic media must be adequate in order to reach the desired supramolecular structure with high fidelity.

Most of the work presented in *Chapter 5* will be published in the following article, which are under preparation:

How Large Can we Build a Cyclic Assembly? Impact of Ring Size on Chelate Cooperativity in Noncovalent Macrocyclizations, C. Montoro-García, M. J. Mayoral, R. Chamorro, D. González-Rodríguez.

In conclusion, a new unconventional and versatile strategy based on molecular self-assembly toward discrete cyclic tetramers has been developed in this Thesis. A large collection of precursors has been synthesized for the construction, *via* Sonogashira coupling reactions, of numerous DNA-based ditopic molecules. These monomers are able to recognize each other and form cyclic tetramers in organic solvents with high fidelity. With these interesting simple molecular building blocks that can be designed at will, countless experiments can be designed for the investigation of their self-assembly under different conditions. As shown in Figure 10, in this Thesis, several topics such as *Supramolecular Equilibria, Symmetric vs Asymmetric H-bonding, Self-sorting Phenomena, Size & Shape Control, Multicomponent Macrocycles and H-bonded Prisms* have been studied. We hope that this novel, bioinspired strategy will further allow to go a step beyond the construction of complex structures from chemically programmed molecules *via* a bottom-up approach.

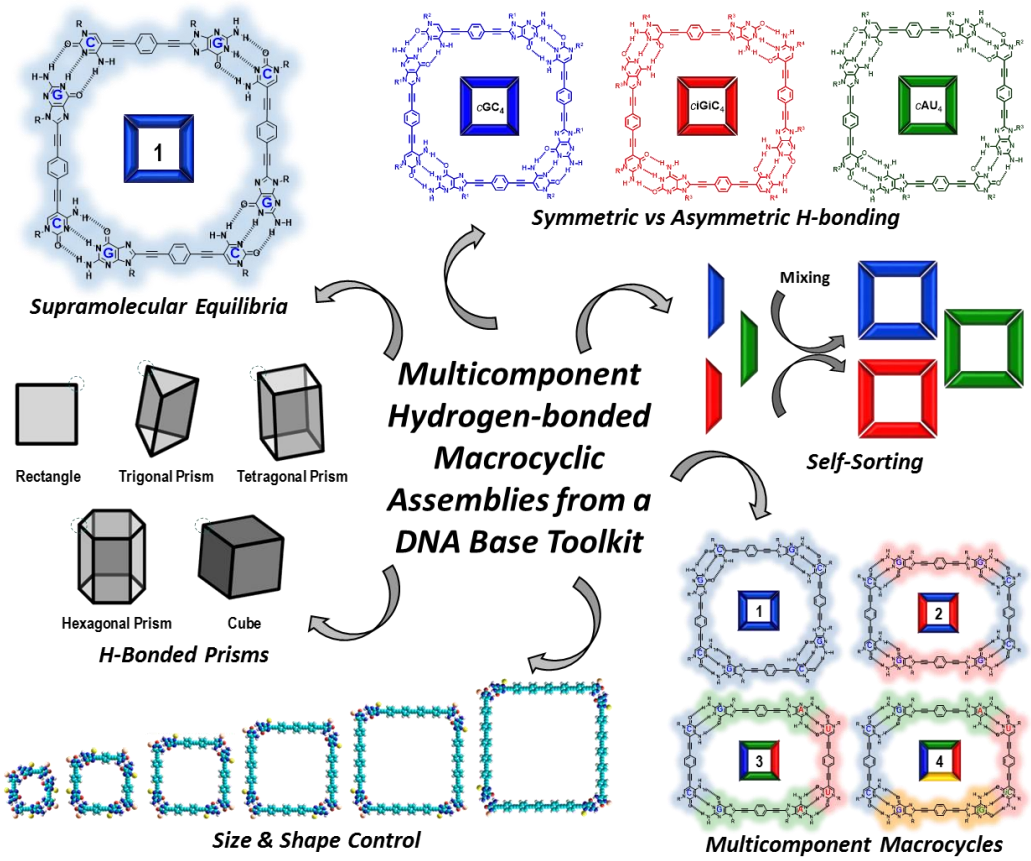


Figure 10. Main topics addressed in *Multicomponent Hydrogen-bonded Macrocyclic Assemblies from a DNA Base Toolkit***.** The general structure of the monomer comprises a rigid unit linearly disubstituted at both termini with nucleosides, so that self-assembly via Watson-Crick pairing between complementary bases in solution results in hydrogen-bonded cyclic tetramers.

Resumen y Conclusiones

Introducción.

El impresionante campo de la ***química supramolecular***, desarrollada en las últimas décadas, ha servido como una herramienta versátil y potente para la construcción de materiales autoensamblados de alta fidelidad a través de la suma de interacciones débiles no covalentes, las cuales se hacen más fuertes mediante efectos cooperativos y multivalentes. Por lo tanto, la química supramolecular se considera como buena opción para construir nanoestructuras a través de la estrategia bottom-up mediante un diseño preciso de moléculas que llevan la información necesaria (tamaño, forma y funcionalidad) para autoasociarse espontáneamente.

Una molécula con más de una posición de autoensamblaje puede asociarse en estructuras lineales (abiertas) o cíclicas (cerradas) en un proceso llamado ***Equilibrio Anillo-Cadena***. Aunque el tamaño de los oligómeros abiertos se puede limitar dentro de un rango concreto, el producto supramolecular es comúnmente una distribución estadística de cadenas de diferente longitud. Por lo tanto, la síntesis de una estructura supramolecular bien definida está enfocada en la obtención de macrociclos cerrados, donde el tamaño está definido por las necesidades geométricas del monómero en cuanto a su dirección de enlace, así como del uso de plantillas que conducen al sistema cíclico. Las especies cíclicas cerradas deben de formarse cuantitativamente debido al aumento de su estabilidad termodinámica, siendo esta muy superior a la suma de las correspondientes interacciones individuales. La ***Cooperatividad Quelato*** es el efecto que causa este incremento de estabilidad, donde la interacción intramolecular esta favorecida frente a la intermolecular, siempre y cuando se cumplan unos requerimientos entálpicos y entrópicos. El incremento de estabilidad cuando se compara un sistema lineal de uno cíclico es dado por el producto K_{inter}·*EM*, donde K_{inter}·es la constante de asociación intermolecular y considera la asociación adicional para formar un ciclo, mientras que *EM*, es un parámetro que cuantifica la Cooperatividad Quelato denominada como ***Molaridad Efectiva***, la cual tiene en cuenta que el último proceso de asociación para formar el ciclo es intramolecular ($EM = K_{intra}/K_{inter}$). Es bien conocido que los monómeros rígidos, teniendo estos una estructura preorganizada que proporciona sistemas no tensionados, son capaces de producir altos valores de *EM* y por lo tanto formar cuantitativamente nanoobjetos discretos autoensamblados.

Objetivo General

En esta Tesis hemos estudiado la formación de sistemas cíclicos en disolución por medio de enlaces de Hidrógeno a partir de monómeros programados químicamente. Mediante el uso de las herramientas que nos proporciona la química supramolecular, el fenómeno cooperativo, y el conocimiento adquirido de las propiedades supramoleculares que poseen los derivados nucleicos del ADN, hemos conseguido un riguroso control de la fidelidad en la escala nanométrica. De este modo, hemos utilizado la alta selectividad y la direccionalidad que nos proporciona el triple enlace de Hidrógeno Watson-Crick en construir sistemas discretos y estables in disolución a partir de monómeros ditópicos que son capaces de autoensamblarse incluso en disolventes polares (Figura 11).

El monómero en cuestión está formado por un ***bloque central*** *rígido y π-conjugado* que está disustituido en ambos finales a través de ***espaciadores*** adecuados con los ***directores*** *de autoensamblaje*, capaces de interactuar mediante enlaces de Hidrógeno.

Los ***directores*** de autoensamblaje, en este caso, son bases nucleicas que forman una familia de tres pares de bases complementarias mediante enlace Watson-Crick: G:C (Guanosina-Citidina), A:U (2-Aminoadenosina-Uridina) y iG:iC (*Iso*guanosina-*Iso*citosina). Cuando se produce este enlace, es importante hacer referencia al ángulo exacto de 90º que se forma entre las posiciones de halogenación selectiva de las bases nucleicas (marcado como **X** en Figura 11), además de que los grupos solubilizadores **R** apuntan siempre hacia el exterior.

Los ***bloques centrales*** tienen que ser rígidos y disustituidos a ambos lados con un ángulo exacto de 180º. Típicamente, suelen ser unidades π-conjugadas que ofrecen planaridad y rigidez al monómero en cuestión. Existe una gran variedad de moléculas que pueden satisfacer estos requerimientos y que poseen diferentes funcionalidades como: TTF, oligómeros conjugados, perilenos, porfirinas ABAB…etc.

Por último, el grupo etinilo ha sido elegido como el ***espaciador*** entre las diferentes unidades que forman el monómero (directores y bloques centrales), ya que satisfacen los requerimientos de rigidez, planaridad y mantienen el ángulo de 180º requerido. Además, posee libertad de rotación, mínimo impedimento estérico y está π-conjugado. Finalmente, presenta una amplia versatilidad química mediante reacciones de acoplación catalizadas por Pd y Pt o mediante reacciones de tipo “Click-Chemistry”.

En general, la estructura lineal (180º) del monómero dinucleosídico junto con el ángulo de 90º impuesto por la interacción Watson-Crick de las bases nucleicas, conducirán a estructuras tetrateméricas compuestas por cuatro monómeros (Figura 11).

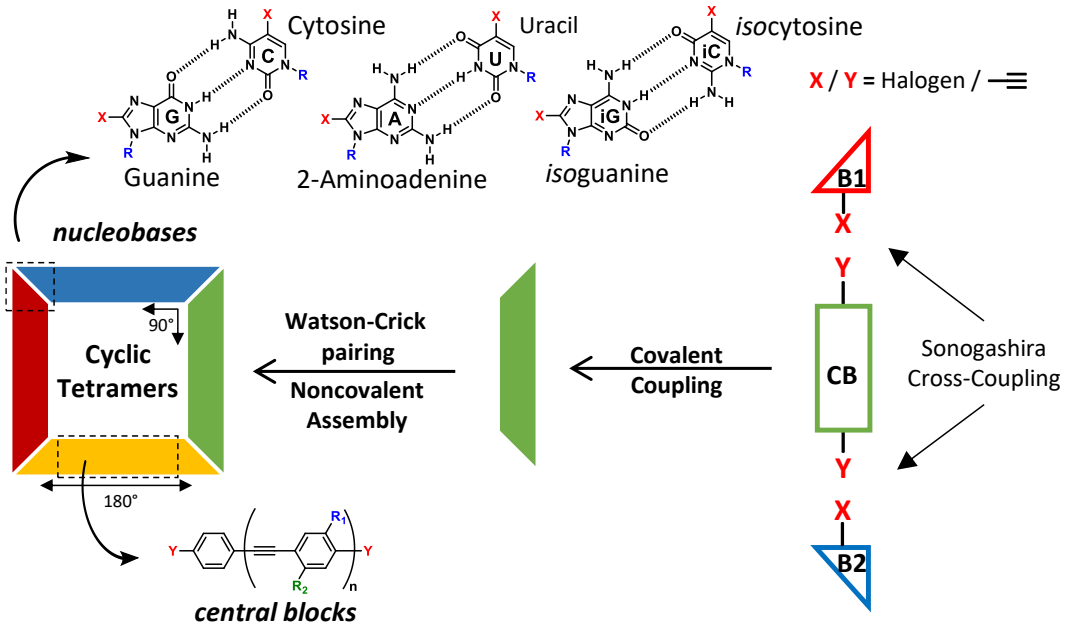


Figura 11. Componentes moleculares y estrategia de autoensamblaje para dar lugar a tetrámeros cíclicos. Bloque Central (**CB**) y Base Nucleicas (**B**).

Capítulo 1. Diseño del Monómero y Síntesis.

Los monómeros finales comparten una estructura común como se ha comentado previamente. En el *Capítulo 1* nos hemos centrado en el diseño y en la síntesis de los diferentes componentes moleculares que los forman. En concreto, hemos desarrollado una nueva y optimizada estrategia para sintetizar una amplia familia de bloques centrales dihalogenados, así como preparar una completa variedad de derivados de bases nucleicas sustituidas con un triple enlace. Una adecuada estrategia para obtener el monómero ditópico a través de reacciones de Sonogashira catalizadas por Paladio fue también abordada en este capítulo. En la Figura 12, se puede observar las diferentes bases nucleicas así como los bloques centrales que se utilizarán como precursores finales para sintetizar los monómeros deseados en cada uno de los capítulos.

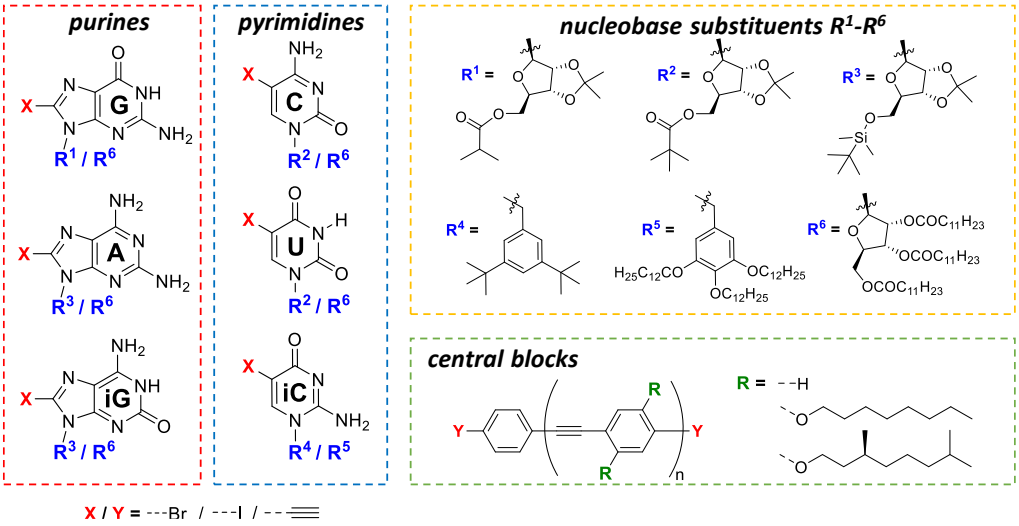


Figura 12. Componentes moleculares considerados y sintetizados en el *Capítulo 1.*

Cinco diferentes moléculas formadas por motivos oligofenileno-vinileno π-conjugados sustituidas con cadenas alquílicas, han sido sintetizadas para completar una colección de siete bloques centrales dihalogenados para ser usados en esta Tesis con distintos propósitos.

Por otro lado, una serie de bases nucleicas lipofílicas han sido preparadas, comprendiendo derivados naturales y no naturales, las cuales fueron sustituidas en las posiciones 5-(pirimidinas) o 8-(purinas) con un átomo de halógeno o con un triple enlace terminal. Estas incluyen la **C**, **iC**, **U** como derivados de pirimidina y **G**, **iG** y **A** como purinas complementarias. El resto ribosa ha sido equipado con diferentes grupos para incrementar la solubilidad en la mayoría de los disolventes orgánicos utilizados, además de prevenir las interacciones π-π entre monómeros. En otro grupo de derivados de bases nucleicas los restos ribosa se funcionalizaron con largas cadenas alquílicas para favorecer aún más la solubilidad en disolventes muy apolares como el tolueno o el tetracloruro de carbono. En otros productos como los derivados de **iC** la ribosa fue reemplazada por dos grupos bencilos diferentes (**R^4** y **R^5** en la Figura 12). La ruta sintética final de los compuestos etinilados ha sido optimizada para cada uno de los casos atendiendo a problemas de reactividad, conveniencia, facilidad de purificación y finalmente por los rendimientos globales. Los resultados finales indican que la elección de la ruta sintética no es trivial y cada derivado requiere un protocolo optimizado.

Capítulo 2. Evaluación de la Constante de Dimerización y Asociación entre Mononucleósidos Lipofílicos.

En el *Capítulo 2* hemos preparado una nueva serie de nucleósidos lipofílicos que están π-conjugados a un fragmento corto de tipo oligofenileno-vinileno. Estas incluyen la **C**, **iC**, **U** como derivados de pirimidina y **G**, **iG** y **A** como purinas complementarias. (Figura 13). Los procesos de dimerización y asociación por enlace de Hidrógeno de las bases complementarias fueron estudiados mediante diferentes técnicas (^{1}H NMR y espectroscopía de absorción), disolventes (tolueno, $CHCl_3$:CCl_4 (2:3), $CHCl_3$, THF y DMF), rangos de concentración, y por último, programas informáticos de ajuste.

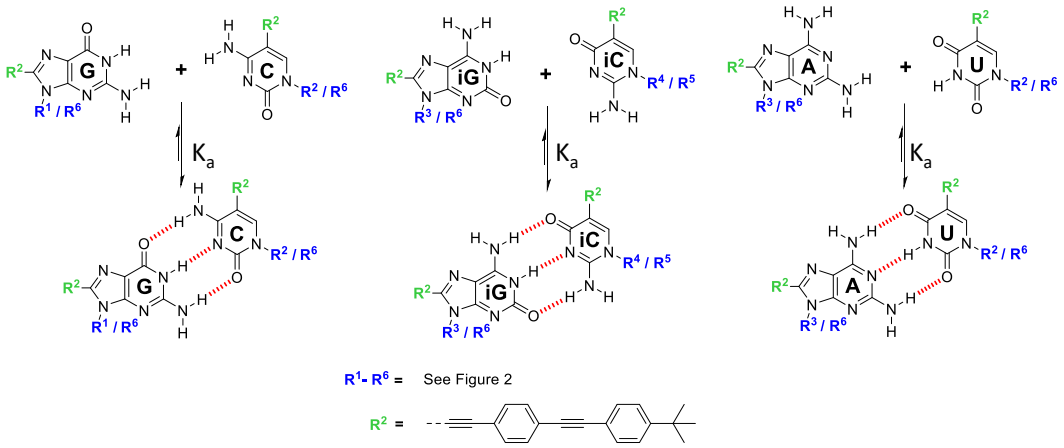


Figura 13. Componentes moleculares considerados y sintetizados en el *Capítulo 2*.

Considerando el $CHCl_3$ (o $CDCl_3$) como disolvente común para comparar la fuerza de enlace, los patrones simétricos de enlace de Hidrógeno *ADA-DAD* (pares de bases **A**-**U**) muestran en una fuerza de enlace en la mezcla 1:1 del orden de 10^2 M^{-1}, mientras que los patrones asimétricos de enlace de Hidrógeno *DDA-AAD* (**G**-**C**, y las nuevas pares de bases **G**-**iC**, **iG**-**C** y **iG**-**iC**) ofrecen unas constantes de asociación en el orden de 10^4 M^{-1}. Este incremento de aproximadamente dos órdenes de magnitud en la estabilidad es conocido en la literatura y es debido al establecimiento de interacciones secundarias de enlace de Hidrógeno en los pares de bases *DDA-AAD,* como se explicó anteriormente en la introducción de la Tesis.

En cambio, las constantes de asociación entre mononucleósidos son muy sensibles a la polaridad del disolvente, donde incrementan un orden de magnitud en disolventes como el tolueno o en mezclas de $CHCl_3$:CCl_4, o descienden cuando el disolvente puede competir fuertemente por el enlace de Hidrógeno como son el THF la DMF.

Este trabajo nos proporciona los valores de la constante de asociación entre pares de bases complementarias en disolventes como tolueno, THF o DMF. Además, incluye los cálculos de K_a para el par no natural **iG**-**iC** y sus posibles combinaciones purina-pyrimidina entre **G**, **C**, **iG** y **iC**. Todas estas combinaciones revelaron similares valores de K_a a consecuencia de su similar patrón de enlace de Hidrógeno *DDA-AAD*.

La mayor parte del trabajo presentado en los *Capítulos 1 y 2* fue publicado en el siguiente artículo científico:

J. Camacho-García, C. Montoro-García, A. M. López-Pérez, N. Bilbao, S. Romero-Pérez, D. González-Rodríguez. "Synthesis and Complementary Self-association of Novel Lipophilic π-Conjugated Nucleoside Oligomers". *Org. Biomol Chem.* **2015**, *13,* 4506-4513 (10.1039/C5OB00098J).

Capítulo 3. Autoensamblaje de Tetrámeros Cíclicos en disolución.

En el *Capítulo 3* nos hemos enfocado en el estudio del proceso de ciclotetramerización en disolución de monómeros ditópicos a través de enlace de Hidrógeno, los cuales llevan pares de bases complementarias (G-C, iG-iC y A-U) en ambos lados. Además, investigamos la fidelidad de este proceso de autoensamblaje, donde la asociación en tetrámeros cíclicos discretos, en lugar de estructuras oligoméricas abiertas u otro tipo de estructuras tensionadas, debería estar altamente favorecida (Figura 14).

El diseño de estos monómeros se basa en la optimización de la *cooperatividad quelato*, obteniendo por lo tanto altos valores de *molaridad efectiva* en el proceso de ciclación. Para este propósito, una serie de monómeros ditópicos, llevando derivados de bases complementarias a cada lado de un resto *p*-fenileno-vinileno linear y rígido (**G1-B1-C1**, **iG1-B1-iC1** y **A1-B1-C1**), han sido preparados. Su proceso de autoensamblaje fue estudiado en diversos disolventes mediante diluciones y variaciones de temperatura, así como a través de experimentos de desnaturalización. Debe subrayarse que, aparte de la excepcional estabilidad termodinámica mostrada por los macrociclos *c***G1-B1-C1**$_4$ / *c***iG1-B1-iC1**$_4$ / *c***A1-B1-U1**$_4$ estudiados en el *Capítulo 3*, los cuales pueden resistir altas diluciones y disolventes polares, constituyen unas estructuras cinéticamente estabilizadas en todo el proceso de autoensamblaje. Los tetrámeros cíclicos se observan siempre sin importar la concentración, la temperatura o las condiciones del disolvente empleado cómo especies en intercambio lento dentro del tiempo de escala de RMN (Figura 15), exhibiendo constantes de intercambio dentro del rango de 1-10 s^{-1}..

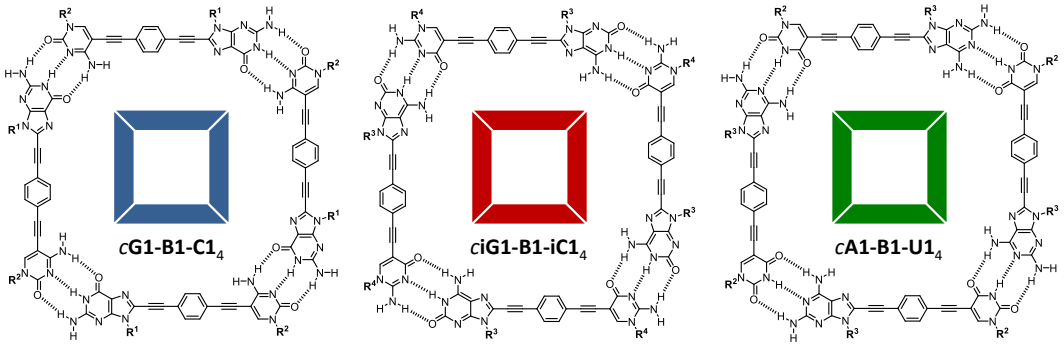


Figura 14. Estructura de los dinucleósidos lipofílicos **G1-B1-C1**, **iG1-B1-iC1** y **A1-B1-C1** considerados en el *Capítulo 3*

Además, en este *Capítulo 3*, hemos examinado el papel que tiene la simetría de los patrones de los enlaces de Hidrógeno en la cooperatividad quelato de los procesos de macrociclación. Hemos sido capaz de separar y analizar independientemente las contribuciones de la fortaleza del enlace de Hidrógeno entre bases nucleicas complementarias, explicadas previamente por el modelo de Jorgensen, y el efecto quelato intrínseco que ejercen en los sistemas cíclicos. Los resultados presentados claramente demuestran que los sistemas cíclicos formados por patrones simétricos *DAD-ADA* de enlace de Hidrógeno, son mucho menos estables que sus homólogos asociados mediante los patrones asimétricos *ADD-DAA* o *DDA-AAD.* Por un lado, el patrón *DAD-ADA* reduce considerablemente el valor entálpico debido a la ausencia de enlaces secundarios atractivos entre enlaces de Hidrógeno vecinales. Por otro lado, la simetría de este patrón introduce la posibilidad de múltiples modos de unión y por lo tanto un mayor número de grados de libertad en los oligómeros lineales, los cuales deben ser perdidos tras la ciclación. Este efecto, de origen entrópico, tiene un gran impacto en el *EM* del sistema, que en nuestro caso se reduce en unos 3 órdenes de magnitud (de *EM* = 10^2-10^3 M para *c***G1-B1-C1**$_4$ / *ci***G1-B1-iC1**$_4$ a *EM* = 0.1-1 M para *c***A1-B1-U1**$_4$).

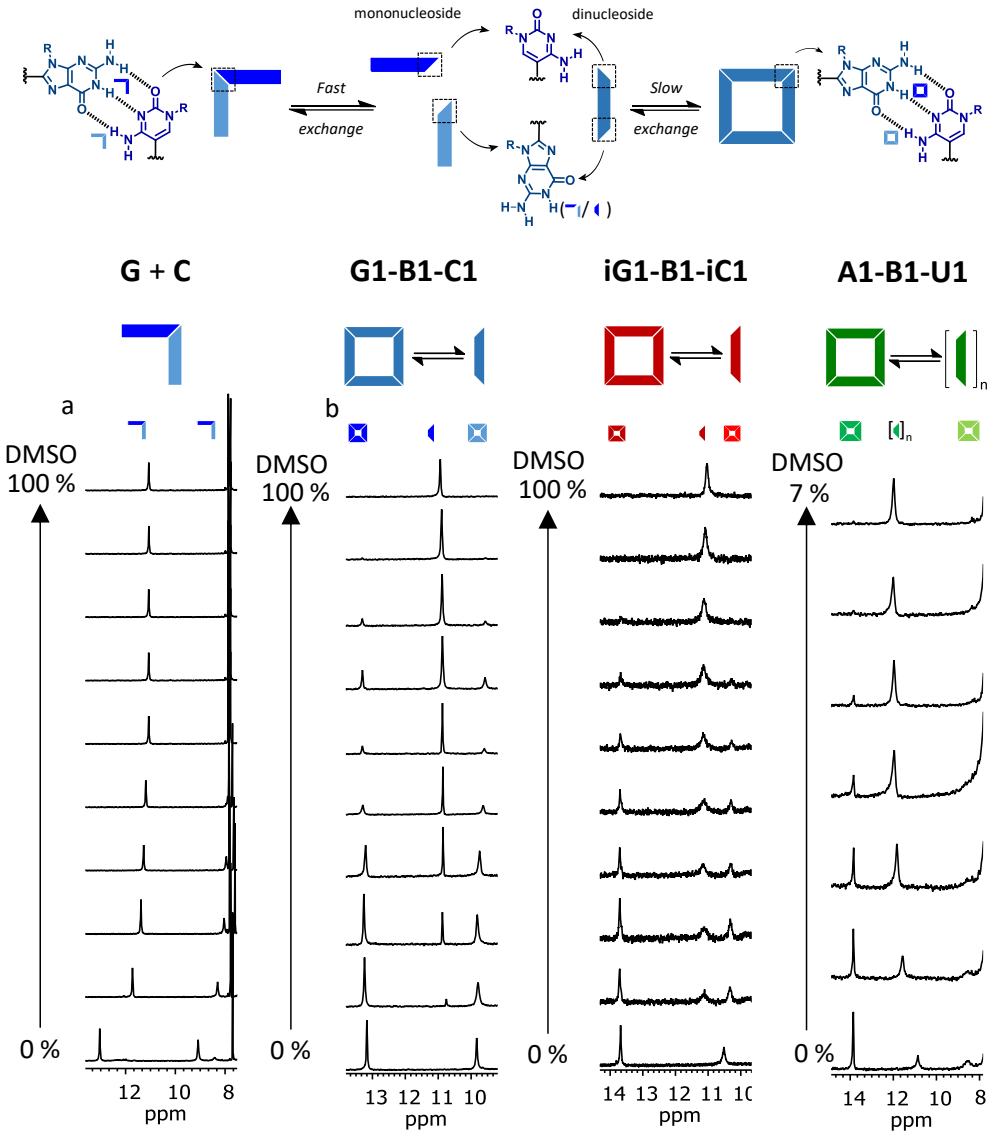


Figura 15. *Experimentos de desnaturalización mediante la adición de DMSO.* Región del espectro de ^{1}H NMR de (**a**) mezcla **G**+**C** 1:1 y (**b**) **G1-B1-C1**, **iG1-B1-iC1** y **A1-B1-U1** cuando la fracción de DMSO-d_6 en $CDCl_3$ va aumentando (C = 1 x 10^{-2} M, K = 298 K en todos los casos). En lo alto de la imagen podemos observar dos diferentes escenarios para el proceso de autoensamblaje entre mononucleósidos y dinucleósidos, los cuales exhiben diferencias termodinámicas y cinéticas. Por un lado, las señales características de **G1**+**C1** en CDCl3 confirman la asociación de este par de bases, mientras que el incremento en pequeñas cantidades de DMSO produce un desplazamiento de las señales a campo alto a través de un equilibrio rápido en el tiempo de medición de ^{1}H NMR entre dimeros y monómeros, debido a que el enlace de Hidrógeno intermolecular esta menos favorecido. Por otro lado, *c***G1-B1-C1**$_4$ muestra una alta resistencia a disolventes polares como *c***iG1-B1-iC1**$_4$, y el tetrámero cíclico persiste incluso después de la adición de más o menos un 80% de DMSO. El tetrámero *c***A1-B1-U1**$_4$ por el contrario, no resiste más de un 7% de DMSO en $CDCl_3$. Además estos experimentos revelaron la presencia de un equilibrio entre monómeros-tetrámeros. En todos los casos según se añade el disolvente polar, la señal correspondiente al monómero crece a expensas de la especie tetramérica. La posición y la forma de señales de **G** (o **U**)-amida y **C** (o **A**)-amina no cambian en posición o forma con la adición, sugiriendo un intercambio muy lento entre especies y un comportamiento "todo o nada", donde el tetrámero es de lejos la especie más estable en disolución o ninguna especie supramolecular puede sobrevivir y la única especie presente es el monómero.

La mayor parte del trabajo presentado en el *Capítulo 3* se publicó en los siguientes artículos:

C. Montoro-García, J. Camacho-García, A. M. López-Pérez, N. Bilbao, S. Romero-Pérez, M. J. Mayoral, D. González-Rodríguez. "High-fidelity Noncovalent Synthesis of Hydrogen-bonded Macrocyclic Assemblies". *Angew. Chem. Int. Ed.* **2015***, 54,* 6780–6784 (10.1002/anie.201501321). (VIP article).

C. Montoro-García, J. Camacho-García, A. M. López-Pérez, M. J. Mayoral, N. Bilbao, D. González-Rodríguez. "Role of the Simmetry of Multipoint Hydrogen Bonding on Chelate Cooperativity in Supramolecular Macrocyclization Processes". *Angew. Chem. Int. Ed.* **2016**, *55,* 223-227 (10.1002/anie.201508854).

Capítulo 4. Fenómeno de Autoclasificación Gobernado por la Cooperatividad Quelato.

Los fenómenos de autoclasificación se han estudiado en el *Capítulo 4*, donde el comportamiento narcisista de la macrociclación de nuestros tres monómeros dinucleosídicos ha sido comparado mediante mezclas binarias de estos con mezclas cuaternarias de los correspondientes mononucleósidos.

Debido a los diferentes patrones de enlace de Hidrógeno Watson-Crick *ADD-DAA* / *DAD-ADA* de las combinaciones **G1-B1-C1** + **A1-B1-U1** y **iG1-B1-iC1** + **A1-B1-U1** esperamos inicialmente su autoclasificación y por lo tanto su autoensamblaje por separado en sus respectivos tetrámeros cíclicos. En el caso de la mezcla **G1-B1-C1** + **iG1-B1-iC1** el patrón de enlace de Hidrógeno es *DDA-AAD* y la predisposición a la autoclasificación no está clara. Como se comentó en el *Capítulo 2*, G puede unirse a iC e iG a C a través la interacción Watson-Crick inversa con una fuerza de enlace comparable con la interacción Watson-Crick G-C e iG-iC. Por ello, la asociación de estos dos dinucleósidos puede conducir a la formación de una mezcla compleja de oligómeros lineales y cíclicos, reduciendo así la fidelidad del proceso de ciclotetramerización.

De hecho, tal como se representa en la figura 16, cuando se mezclan los monómeros **G1-B1-C1** y **A1-B1-U1**, las señales características de sus respectivos tetrámeros mono-componente aparecieron por separado, lo cual es indicativo de su autoclasificación. Sin embargo, cuando se mezclan los monómeros **G1-B1-C1** e **iG1-B1-iC1**, las señales de sus macrociclos correspondientes se encuentran de nuevo separadas en los espectros de ^{1}H NMR y no hay signos de ningún otro tipo de especie asociada o no asociada. Por otro lado, los experimentos NOESY de ambas mezclas de nucleósidos confirman además que A sólo se une a U, G sólo se une a C, e iG sólo se une a iC.

Estos resultados confirman claramente el comportamiento narcisista de autoclasificación entre los monómeros **G1-B1-C1** y **A1-B1-U1** que era un comportamiento esperado, y entre **G1-B1-C1** y **iG1-B1-iC1**. Sin embargo, en el último caso, los fenómenos de autoclasificación no se rigen por el patrón de enlace de enlace de Hidrógeno, sino más bien por la alta *cooperatividad quelato* mostrada por sus respectivos tetrámeros cíclicos, como determina el *Capítulo 3*, que predice que la estructura autoensamblada más estable es la más pequeña posible y la menos tensionada. Los tetrámeros cíclicos solo se pueden producir cuando al formarse la interacción Watson-Crick, las posiciones 8-purina y 5-pirimidina adquieren un ángulo de 90°. Esta nueva fuerza impulsora que dirige los procesos de autoclasificación en química supramolecular se tendrán en cuenta para desarrollar conjuntos de mezclas más sofisticados.

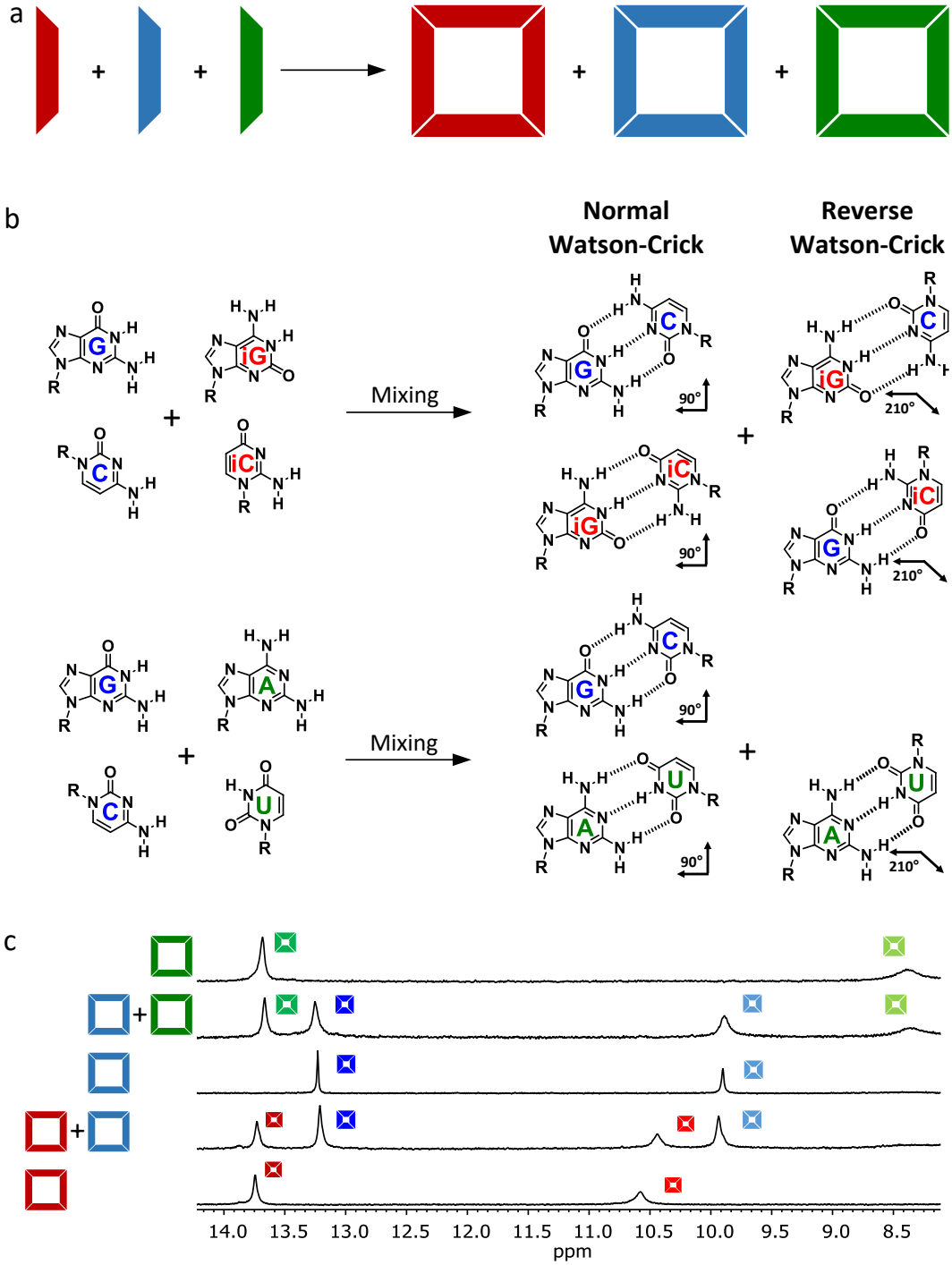


Figura 16. *Fenómeno de autoclasificación de los monómeros ditópicos.* (**a**) Diferentes parejas de bases unidas mediante enlace Watson-Crick estudiadas en el *Capítulo 4* y (**b**) región del espectro de ^{1}H NMR mostrando la señal unida por enlace de Hidrógeno de G-amida (H^1; azul oscuro) y C-amina (H^2; azul claro), U-imida (H^1; verde oscuro) and A-amina (H^2; verde claro), iG-amida (H^1; rojo oscuro) e iC-amina (H^2; rojo claro) de **G1-B1-C1, A1-B1-U1**, e **iG1-B1-iC1** y sus respectivas mezclas.

La mayor parte del trabajo presentado en el *Capítulo 4* será publicado en el siguiente artículo que actualmente está siendo preparado:

"Self-sorting Phenomena Governed by Chelate Cooperativity", C. Montoro-García, D. Serrano-Molina, M. J. Mayoral, D. González-Rodríguez, *en preparación*

Capítulo 5. Impacto del Tamaño del Anillo en la Cooperatividad Quelato en Macrociclaciones No Covalentes.

Una vez estudiado el proceso de macrociclación y autoclasificación de los monómeros **G1-B1-C1**, **iG1-B1-iC1** y **A1-B1-U1**, el *Capítulo 5* está centra en el estudio del impacto que sufre la cooperatividad quelato cuando la longitud del bloque central es elongado o acortado. Estas modificaciones varían desde un enlace sencillo C-C hasta una estructura penta-(fenileno-vinileno) π-conjugada (Figura 17). Todos estos bloques centrales están sustituidos a ambos lados por la misma pareja de bases G-C, ya que como se explicó previamente, el motivo asimétrico del patrón de enlace de Hidrógeno aumenta la cooperatividad quelato y es por lo tanto la mejor manera para comparar la estabilidad que cada uno de los bloques centrales confiere a tetrámero cíclico.

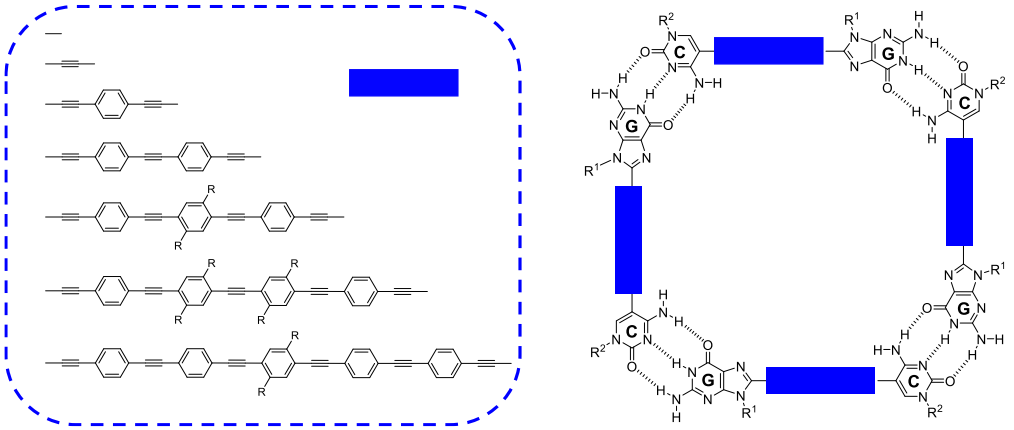


Figure 17. Bloques centrales con diferentes longitudes entre el par de bases G-C estudiados en el *Capítulo 5*.

Los estudios y análisis de autoensamblaje de los diversos monómeros a través de mediciones dependientes de la concentración y de la temperatura, así como mediante experimentos de desnaturalización con un codisolvente polar (DMSO) o el mononucleósido **C** complementario, demuestran claramente que el impacto de cada bloque central colocado entre las bases nucleicas complementarias en la cooperatividad quelato se manifiesta de dos maneras diferentes.

En primer lugar, cuando el bloque central se hace más grande que el bloque clásico de *p*-fenileno-dietileno estudiado en el *Capítulo 3*, con longitudes de monómeros que van desde 2.1 a 4.7 nm, hay una pérdida notable del valor de la molaridad efectiva que puede abarcar 5 órdenes de magnitud. Se observó que este efecto sólo se debe a factores entrópicos, mientras que el componente entálpico en el proceso de ciclotetramerización permanece constante para todas las moléculas G-C, ya que la geometría del monómero y la interacción de unión son las mismas. Nuestra explicación es que, a medida que aumenta el número de enlaces *σ*, la flexión y torsión del bloque central oligofenileno-vinileno aumenta, alcanzando grados adicionales de libertad que deben perderse para la ciclación (Figure 18). ΔS y Ln *EM* parecen

que siguen una relación linear con el número de enlaces σ siendo independiente del disolvente empleado. La extrapolación de estas tendencias proporciona una estimación de cuán grande podemos construir un tetrámero cíclico usando este tipo de monómeros. Cualquier cambio en este tipo de estructura oligofenileno-etinileno puede producir una desviación importante de esta tendencia como demuestran los monómeros equipados con espaciadores bifenilo.

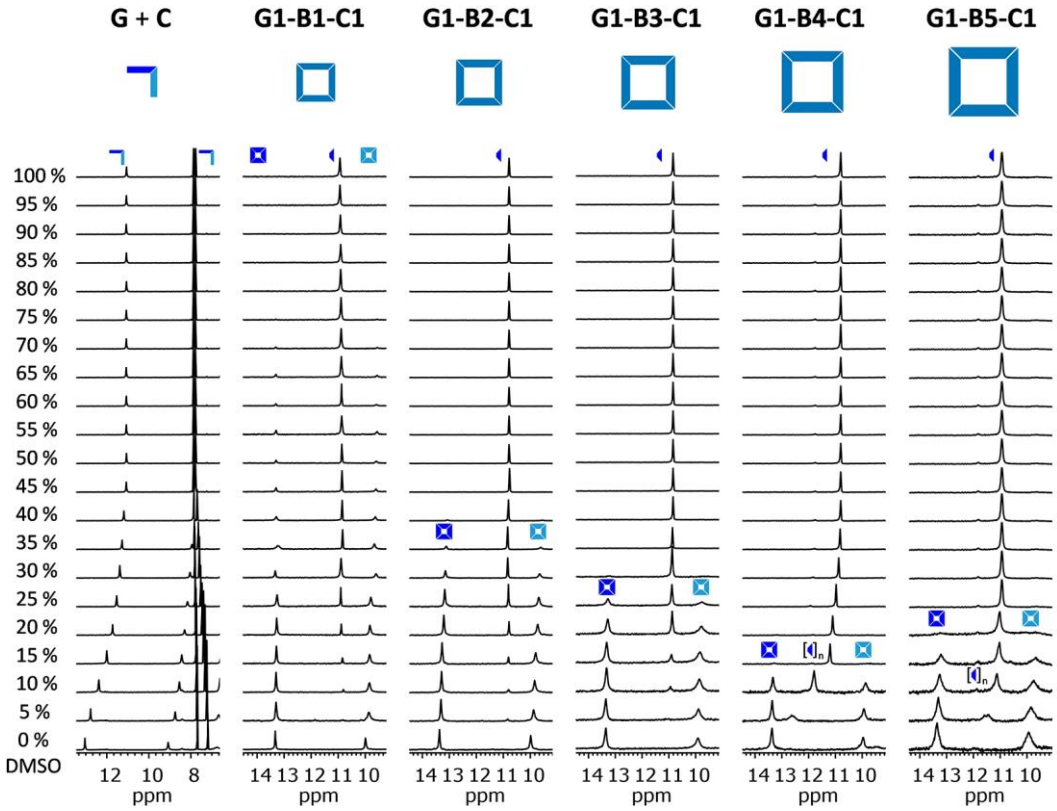


Figura 18. *Experimentos de Desnaturalización Mediante la Adición de DMSO***.** Región del espectro de ^{1}H NMR que muestra la señal de protón unida por enlace de Hidrógeno de G-amida (H^1; azul oscuro) y C-amina (H^2; azul claro) de la mezcla 1:1 **G+C** y de los dinucleósidos **G1-B1-C1** – **G1-B5-C1** cuando la fracción de DMSO-d_6 se incrementa en mezclas $CDCl_3$-DMSO-d_6 (C = 1.0x10^{-2} M y T = 298 K. en todos los casos).

En segundo lugar, cuando el bloque central se hace más corto, como un triple enlace o un solo enlace C-C entre bases nucleicas, otros efectos se vuelven importantes. El impedimento estérico es uno de ellos y tiene un papel importante para alcanzar la deseada conformación "*syn*" entre bases nucleicas, la cual es necesaria para obtener un tetrámero cíclico. El volumen y la dirección en la que los sustituyentes de las bases nucleicas están apuntando es crucial (Figura 19) para decidir si estas moléculas "más cortas" se ensamblarán como polímeros supramoleculares viscosos en numerosos disolventes o en asociaciones discretas cerradas.

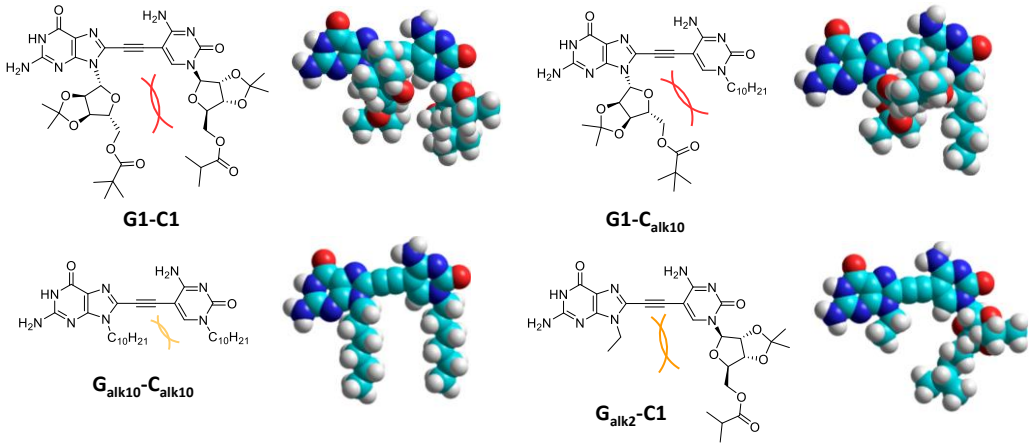


Figura 19. Impedimento estérico y modelo molecular (derecha) de los diferentes monómeros unidos por grupos etinilo sintetizados en el *Capítulo 5*. El color indica el grado de impedimento estérico, rojo (alto) y naranja (moderado).

Nuestras conclusiones podrían extenderse en principio a muchos ciclos supramoleculares o cajas moleculares en donde tamaño del monómero se alarga o acorta. Por lo tanto, el diseño previo de los bloques centrales y los sustituyentes que confieren solubilidad en medios orgánicos deben ser adecuados para alcanzar la estructura supramolecular deseada con una alta fidelidad.

La mayor parte del trabajo presentado en el *Capítulo 5* será publicado en el siguiente artículo que está siendo preparado actualmente:

How Large Can we Build a Cyclic Assembly? Impact of Ring Size on Chelate Cooperativity in Noncovalent Macrocyclizations, C. Montoro-García, M. J. Mayoral, R. Chamorro, D. González-Rodríguez. *En preparación.*

Como conclusión, una nueva estrategia no convencional y versátil basada en el autoensamblaje para la formación de tetrámeros cíclicos discretos ha sido desarrollada en esta Tesis. Se prepararon una amplia colección de precursores moleculares para la obtención de monómeros ditópicos basados en el ADN que se unieron posteriormente a través de reacciones de Sonogashira. Estos monómeros sintetizados son capaces de reconocer a sus homólogos y formar tetrámeros cíclicos con una alta fidelidad en una amplia variedad de disolventes orgánicos. Con estos monómeros, los cuales se pueden sintetizar a nuestra voluntad, se pueden diseñar una amplia cantidad de experimentos para estudiar su proceso de autoensamblaje bajo diferentes condiciones. Como se muestra en la Figura 20, los principales temas de esta Tesis han sido: *el Equilibrio Supramolecular, la influencia del Patrón de Simetría en el Enlace de Hidrógeno, los fenómenos de Auto-Clasificación, el Control de Tamaño y Forma de los Tetrámeros, los Macrociclos Multicomponentes y finalmente la Formación de Prismas Moleculares en disolución.* Esperamos que esta nueva estrategia, inspirada en el mundo biológico, permita en un futuro próximo la construcción de estructuras más complejas a partir de moléculas programadas químicamente a través de una estrategia bottom-up.

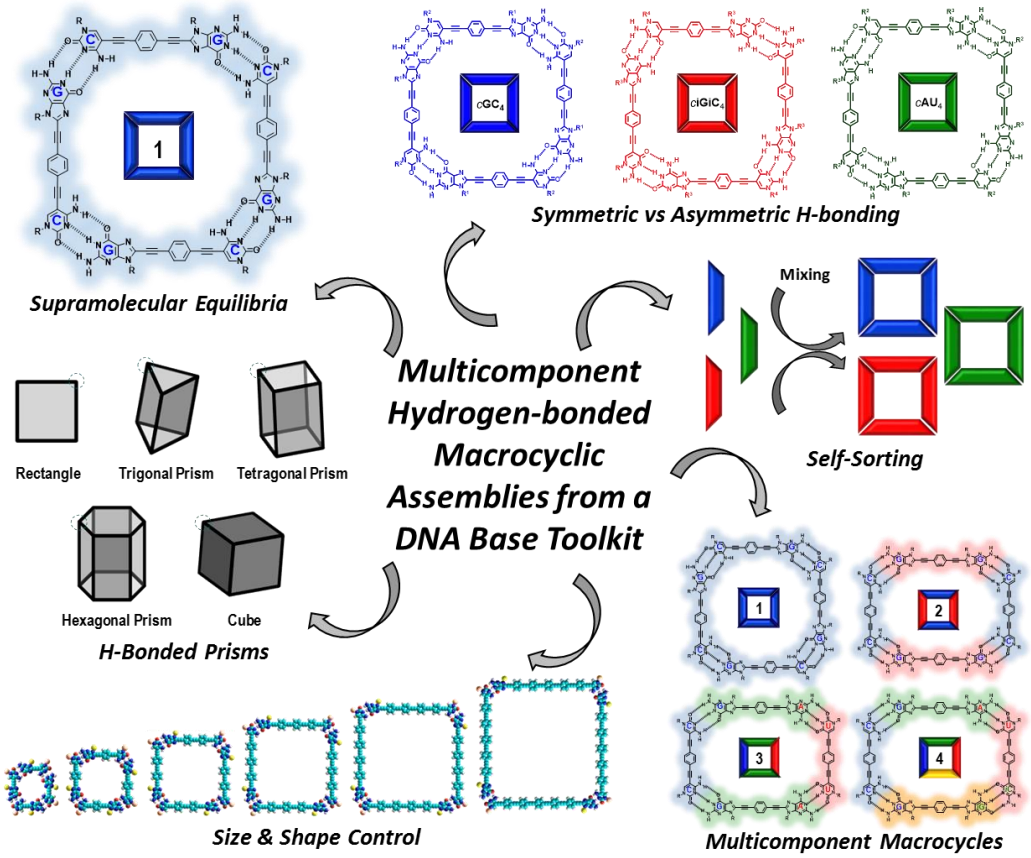


Figure 20. Principales temas de la Tesis Autoensamblaje *Multicomponent Hydrogen-bonded Macrocyclic Assemblies from a DNA Base Toolkit*. La estructura general del monómero comprende un bloque central rígido disustituido a ambos lados con derivados de bases nucleicas, por lo que podrán autoensamblarse mediante enlaces de Hidrógeno para la formación de tetrámeros cíclicos en disolución.