Self-Assembly

Deutsche Ausgabe: DOI: 10.1002/ange.201508854
Internationale Ausgabe: DOI: 10.1002/anie.201508854

Role of the Symmetry of Multipoint Hydrogen Bonding on Chelate Cooperativity in Supramolecular Macrocyclization Processes

*Carlos Montoro-García, Jorge Camacho-García, Ana M. López-Pérez, María J. Mayoral, Nerea Bilbao, and David González-Rodríguez**

Abstract: *Herein, we analyze the intrinsic chelate effect that multipoint H-bonding patterns exert on the overall energy of dinucleoside cyclic systems. Our results indicate that the chelate effect is regulated by the symmetry of the H-bonding pattern, and that the effective molarity is reduced by about three orders of magnitude when going from the unsymmetric ADD–DAA or DDA–AAD patterns to the symmetric DAD–ADA pattern.*

The supramolecular synthesis[1] of complex nanostructures with a precision analogous to that found in the natural world requires an understanding not only of the noncovalent interactions involved,[2] but also of cooperative and multivalent phenomena that may arise between the individual constituents, since the control of structure and monodispersity depends largely on this issue.[3] A molecule with more than one binding site may assemble into linear (open) or cyclic (closed) structures. Although the size of linear oligomers can sometimes be limited within a certain range, the supramolecular product is commonly a statistical distribution of chain lengths.[4] Therefore, the synthesis of discrete supramolecular structures has normally been focused on closed (multi)-macrocyclic systems, in which size and structure are dictated by the geometric requirements of the monomer and the binding interaction.[5] The effect that causes the quantitative formation of a particular ring-closed species is defined as chelate cooperativity and stems from the fact that an intramolecular interaction is favored over an intermolecular interaction, providing that a series of conditions of enthalpic and entropic origin are met.[3] The increased stability of a cyclic oligomer relative to that of the corresponding linear oligomer is given by the product $K_{inter}{\cdot}EM$, in which K_{inter} is the intermolecular binding constant and considers the additional association to form the macrocyclic ring, and *EM*, the key parameter in the quantification of chelate cooperativity, stands for effective molarity and takes into account that this last binding event is intramolecular ($=K_{intra}/K_{inter}$).[6]

In this context, multipoint H-bonding motifs, constituted by an array of vicinal H-bonding donor (D) and acceptor (A) groups, arise as a relevant noncovalent interaction increasingly used to produce not only discrete cyclic assemblies, but also supramolecular polymers and functional materials.[7] The nucleobases are a relevant example,[8] and DNA itself, composed of combinations of unsymmetric ADD–DAA guanine–cytosine and symmetric DA–AD adenine–thymine H-bonded Watson–Crick pairs, can be regarded as the biological stereotype of a closed assembly. The relative strength of these multipoint H-bonding interactions is now well-understood since the investigations described by Jorgensen and Pranata in 1990.[9] Their interpretation takes into account secondary electrostatic interactions between contiguous centers to explain the trend in the association constants of, for example, triply H- bonded pairs: DDD–AAA > ADD–DAA > DAD–ADA. However, the intrinsic influence of the H-bonding pattern on *EM*, and hence on the chelate cooperativity of a cyclization process, has never been addressed, and was the main focus of this study. We have compared the thermodynamics of the self-assembly of three related monomers (**GC**, **iGiC**, **AU**) into their respective cyclic tetramers ($c\mathbf{GC}_4$, $c\mathbf{iGiC}_4$, $c\mathbf{AU}_4$; Figure 1). Our results indicate a huge effect of the symmetry of the binding interaction on the magnitude of *EM*, and may thus be highly valuable in

Table 1: Cyclotetramerization constants (K_T), reference intermolecular association constants (K_{ref}), and effective molarities (*EM*) obtained for **GC**/**iGiC**/**AU** from different experiments.

M	Solvent	K_T [M^{-3}]	K_{ref}[a] [M^{-1}]	*EM* [M]
GC	DMF	$2.3\pm0.8\times10^{5}$[b]	5.7 ± 0.3	218
	THF	$9.1\pm4.0\times10^{14}$[c]	$1.5\pm0.1\times10^{3}$	180
		$3.7\pm0.3\times10^{15}$[d]		730
	$CHCl_3$	$5.6\pm3.1\times10^{20}$[e]	$2.8\pm0.3\times10^{4}$	910
		$5.0\pm0.1\times10^{20}$[f]		813
iGiC	DMF	$3.4\pm1.9\times10^{5}$[b]	6.1 ± 0.8	246
	THF	$3.7\pm1.2\times10^{15}$[c]	$1.7\pm0.6\times10^{3}$	463
		$2.2\pm0.5\times10^{15}$[d]		294
	$CHCl_3$	$3.3\pm0.4\times10^{20}$[f]	$3.2\pm0.5\times10^{4}$	314
AU	$CHCl_3$		$2.5\pm0.4\times10^{2}$	0.10[g]
	$CHCl_3/CCl_4$ (2:3)	$9.4\pm0.3\times10^{11}$[c]	$2.0\pm0.4\times10^{3}$	0.06
		$2.8\pm0.2\times10^{11}$[d]		0.02
	$CHCl_3$/acetone (5:1)	$7.2\pm1.6\times10^{6}$[h]	$0.9\pm0.6\times10^{2}$	0.11

[a] Determined from titration experiments with the mononucleosides: **G** + **C**, **iG** + **iC**, **A** + **U**.[12] [b] Determined from ^{1}H NMR dilution experiments (see Figure S9). [c] Determined from UV/Vis dilution experiments (see Figure S13). [d] Determined from temperature-dependent experiments (see Figure S14). [e] Determined from ^{1}H NMR competition experiments (see Figure S15). [f] Determined from fluorescence competition experiments (see Figure S16). [g] Estimated from the fitting of the ^{1}H NMR dilution data (see Figure S1 and Figure 2c). [h] Determined from ^{1}H NMR dilution experiments (see Figure S7B). DMF = *N*,*N*-dimethylformamide, THF = tetrahydrofuran.

[*] C. Montoro-García, J. Camacho-García, Dr. A. M. López-Pérez, Dr. M. J. Mayoral, N. Bilbao, Dr. D. González-Rodríguez
Nanostructured Molecular Systems and Materials Group
Departamento de Química Orgánica
Facultad de Ciencias, Universidad Autónoma de Madrid
28049 Madrid (Spain)
E-mail: david.gonzalez.rodriguez@uam.es

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201508854.

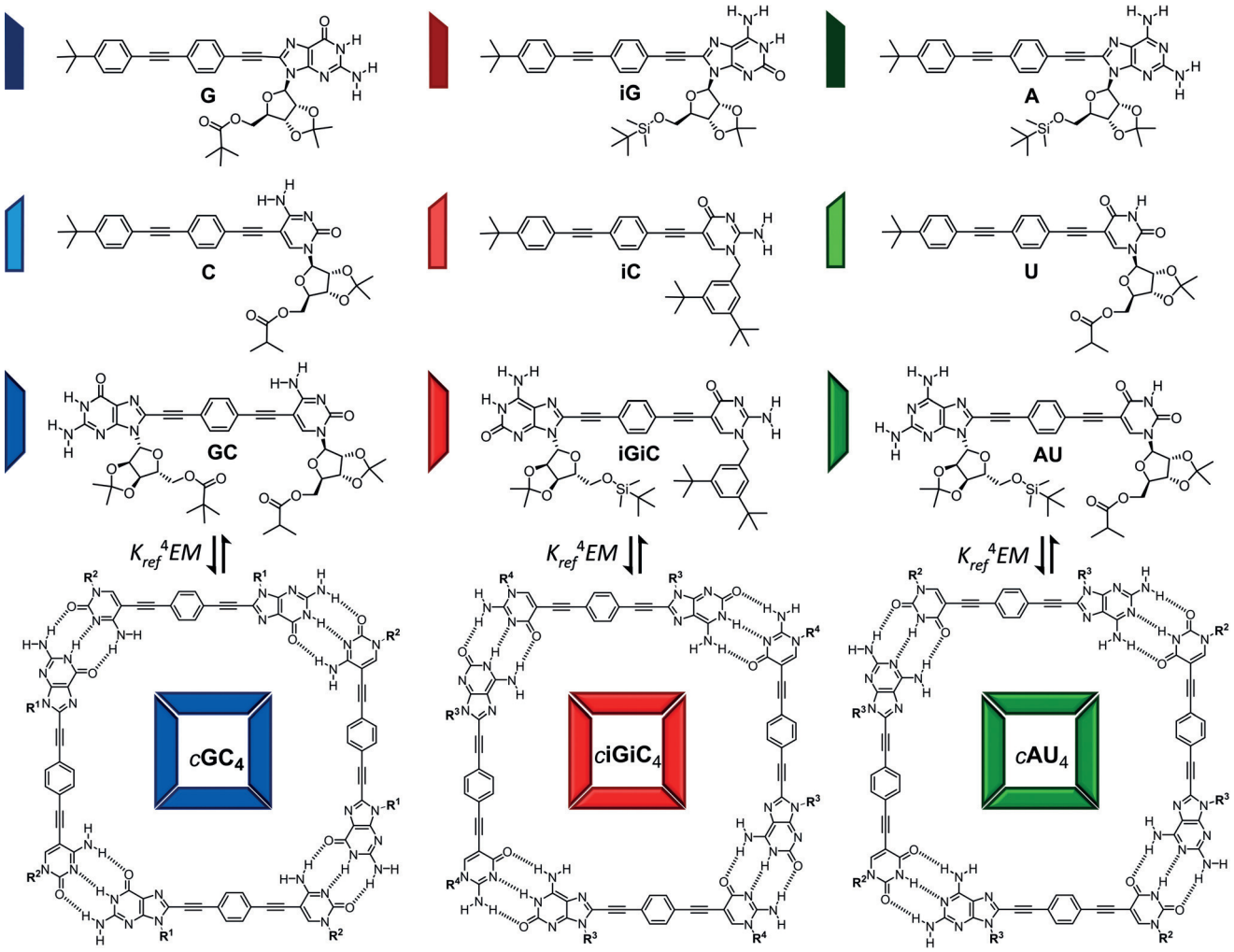

Figure 1. Structure of lipophilic dinucleoside monomers **GC**, **iGiC**, and **AU**, reference mononucleoside compounds **G**, **C**, **iG**, **iC**, **A**, and **U**, and the three cyclic tetramers formed in solution: *c***GC**$_4$ (ADD–DAA), *c***iGiC**$_4$ (DDA–AAD) and *c***AU**$_4$ (DAD–ADA).

the future design of suitable monomers that lead quantitatively to either discrete closed assemblies or, at the other extreme, to supramolecular polymers.

The self-assembly of monomers **GC**, **iGiC**, and **AU** was analyzed in different solvents by a number of concentration- and temperature-dependent spectroscopic methods (1D and 2D ^{1}H NMR spectroscopy, as well as absorption, emission, and CD spectroscopy). These experiments were described in our previous work[10] and the combined results for the 3 monomers are detailed in the Supporting Information of this manuscript (see Figures S1–S16).[11] The equilibrium constants of the cyclotetramerization processes (K_T) could be obtained in some cases and are collected in Table 1, together with the reference association constants for the interaction between the complementary mononucleosides (K_{ref}). The whole set of results clearly demonstrate that **AU** forms considerably less stable cyclic tetramers than **GC** or **iGiC**. On the other hand, **GC** or **iGiC** show comparable qualitative and quantitative association behavior.

This stability trend was expected, since the individual **A**:**U** binding constant ($K_{ref} \approx 2.5 \times 10^2\,M^{-1}$) is typically about two orders of magnitude lower than that for **G**:**C** or **iG**:**iC** ($K_{ref} \approx 3 \times 10^4\,M^{-1}$) in $CHCl_3$.[12] This can be explained by 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**) pair.[9] However, our experimental results (Table 1) suggest that the decrease in stability of the corresponding cyclic tetramer is actually much larger. As explained below, our results indicate that the K_T value for **AU** is not only reduced by a decrease in K_{ref}, but also by a substantial decrease in the magnitude of *EM*.

To evaluate this hypothesis, we performed competition experiments in which the corresponding complementary pyrimidine mononucleoside (**C**/**iC**/**U**) was gradually added to a solution of the associated tetramers (*c***GC**$_4$/*c***iGiC**$_4$/*c***AU**$_4$; see Figures S15 and S16). The titration process was monitored in solvents of low polarity by two different techniques: ^{1}H NMR spectroscopy at high concentrations (ca. 10^{-2} M) and emission spectroscopy at relatively low concentrations (ca. 10^{-4} M). Our results show that whereas *c***GC**$_4$/*c***iGiC**$_4$ can resist up to 60 equivalents of the **C**/**iC** mononucleoside, *c***AU**$_4$ was fully dissociated after the addition of about 3 equivalents of **U**, regardless of the solvent system employed, thus indicating

a much weaker chelate effect in $c\mathbf{AU}_4$ (Figure 2a). These experiments, in which the intramolecular and intermolecular base-pair binding events are made to compete, constitute the most appropriate way to detach the intrinsic contribution of the chelate cooperativity from the overall energy of the system. As a matter of fact, the *EM* of the system can be inversely related to the competition equilibrium constant (K_C) as: $EM = 1/K_C$.[11]

The *EM* values for each cyclotetramerization process (Table 1) were calculated from the K_T values and the corresponding K_{ref} constants obtained in the different experiments[10,12] by use of the relationship: $EM = K_T/K_{ref}^4$. These K_{ref} and *EM* values were also used to simulate speciation curves for each dinucleoside in different solvent systems (Figure 2b,c). These curves, which relate the concentration of each supramolecular species to the total concentration, reproduce quite satisfactorily the dissociation trends observed for $c\mathbf{GC}_4/c\mathbf{iGiC}_4/c\mathbf{AU}_4$ in dilution experiments.[11] In all cases, $c\mathbf{GC}_4$ and $c\mathbf{iGiC}_4$ fulfill the condition $K_{ref}{\cdot}EM > 185{\cdot}n$ (in which n is the number of monomers in the cycle; $n = 4$), as defined by Ercolani for complete cycle assembly.[6] However, the concentration range in which 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. On the other hand, *EM* values had to be set between 0.05 and 0.1 to reproduce appropriately the dilution trends of $c\mathbf{AU}_4$ in three solvent systems. They are thus more than three orders of magnitude lower than those of the **GC/iGiC** cyclization process. As a result, and in line with our experimental observations, the condition $K_{ref}{\cdot}EM > 185{\cdot}n$ is hardly fulfilled by **AU** even in the highly apolar $CHCl_3/CCl_4$ (2:3) mixture, in which K_{ref} is enhanced.

All monomers share a common, rigid structure that was designed to produce cyclic square-shaped assemblies devoid of strain and with minimal conformational entropy loss. These properties were made possible by the 90° angle that the 8-purine and 5-pyrimidine positions adopt upon Watson–Crick complementary base pairing and by the use a rigid central block to connect the bases with only four rotatable linear π-conjugated bonds. Rotation about these bonds can produce different conformations in which the Watson–Crick edges alternate between *syn* and *anti* relative arrangements (Figure 3a). However, cycle formation demands that the Watson–Crick edges are in a *syn* relative conformation. This is a degree of freedom that is lost when comparing cyclic and open **GC**, **iGiC**, and **AU** oligomers, and must contribute to a reduction, of entropic origin, in the maximum attainable *EM* of the cyclic system. Now, the **AU** monomer, which contain complementary nucleosides that pair with a symmetric DAD–ADA H-bonding pattern, have the additional possibility to self-assemble through either Watson–Crick or reverse Watson–Crick interactions (Figure 3a). Each binding mode provides a different association angle (90° and 210°), and their relative energy is assumed to be comparable, as previous studies with the adenine–thymine pair have demonstrated.[13,14] This property introduces additional degrees of freedom, not available in the unsymmetric ADD–DAA or DDA–AAD patterns, that allow the linear $\mathbf{AU}_n$ oligomers to access

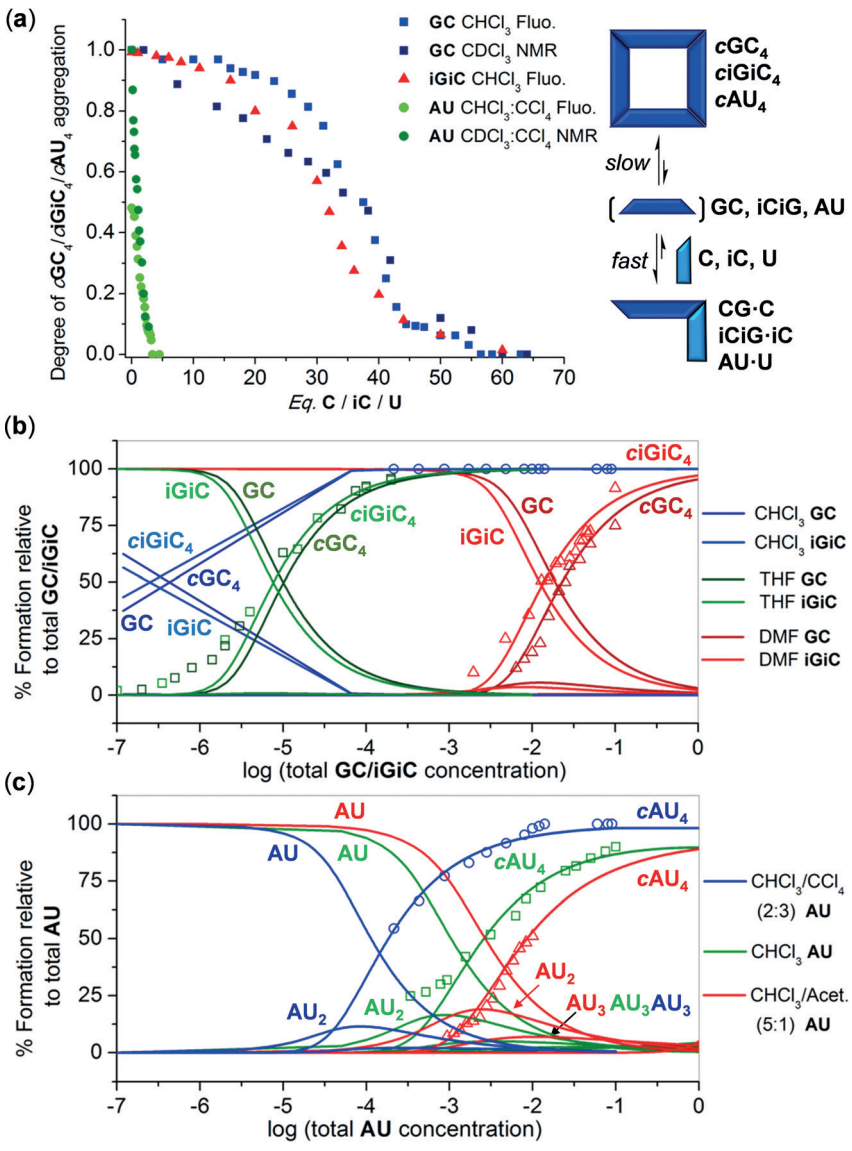

Figure 2. a) Competition experiments. Plots of the degree of $c\mathbf{GC}_4$, $c\mathbf{iGiC}_4$, and $c\mathbf{AU}_4$ association, as measured by either ^{1}H NMR or fluorescence spectroscopy, as a function of the equivalents of **C**/**iC**/**U** added. b,c) Simulated speciation curves (lines) and experimental dilution data (circles/triangles/squares) indicating the degree of $c\mathbf{M}_4$ association of b) **GC**/**iGiC** in $CHCl_3$ (see Figure S1), THF (see Figure S13A,B), and DMF (see Figure S9), and c) **AU** in $CHCl_3/CCl_4$ (2:3; see Figure S13C), $CHCl_3$ (see Figure S1), and $CHCl_3$/acetone (5:1; see Figure S7B).

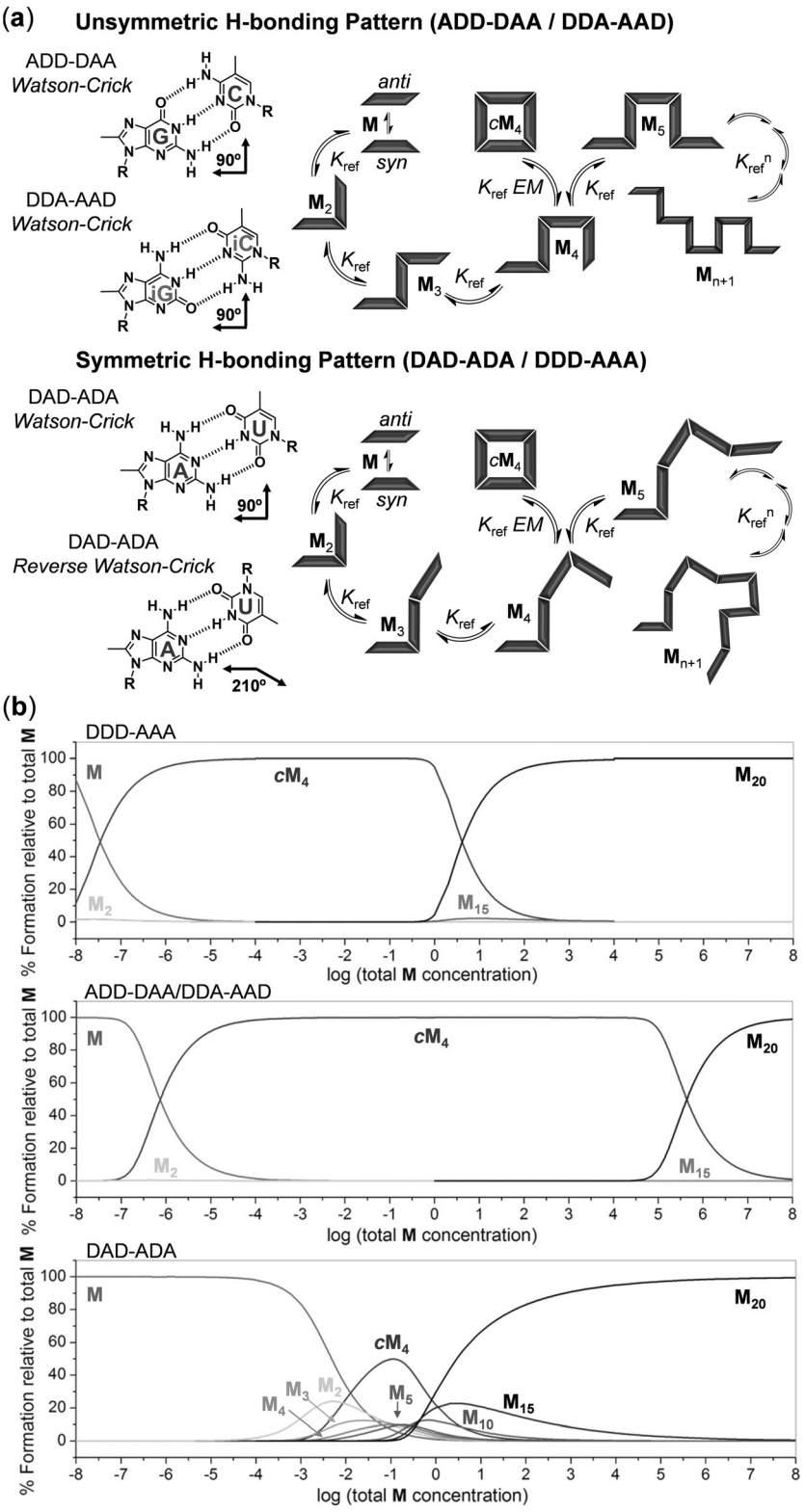

Figure 3. a) Unsymmetric versus symmetric H-bonding patterns. b) Simulated speciation curves for the H-bonded self-assembly of hypothetical DDD–AAA ($K_{ref}=10^6\,M^{-1}$; $EM=0.05\,M$), ADD–DAA/DDA–AAD ($K_{ref}=10^4\,M^{-1}$; $EM=500\,M$), and DAD–ADA ($K_{ref}=10^2\,M^{-1}$; $EM=0.05\,M$) ditopic monomers with identical geometrical features to those examined in this study. For the sake of clarity, only a few supramolecular species are represented within the 10^{-8}–$10^8\,M$ concentration range: **M**, $\mathbf{M}_2$, $\mathbf{M}_3$, $\mathbf{M}_4$, $c\mathbf{M}_4$, $\mathbf{M}_5$, and, as examples of higher-order H-bonded linear oligomers: $\mathbf{M}_{10}$, $\mathbf{M}_{15}$, and $\mathbf{M}_{20}$.

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, we assume that the entropy loss associated with cyclization becomes much larger for $c\mathbf{AU}_4$ than for $c\mathbf{GC}_4$ or $c\mathbf{iGiC}_4$, thus resulting in a supplementary and notable reduction of the *EM* values.[15]

It would have been highly interesting to compare the cyclization process of our **AU**, **GC**, and **iGiC** molecules with that of a related dinucleoside with complementary DDD–AAA base pairs. However, a purine–pyrimidine couple with such a H-bonding pattern that would maintain the same geometric requirements is simply not available. We propose, given the conclusions drawn from this study, that such a cyclic tetramer bound by symmetric DDD–AAA pairs would have both a high K_{ref} value (ca. 10^5–$10^6\,M$ in $CHCl_3$, as reported previously)[7a] and a low *EM* (0.01–0.1, comparable to that of $c\mathbf{AU}_4$), owing to the possibility of binding with two different angles. From these values, we simulated in Figure 3b the speciation curves of this hypothetical symmetric DDD–AAA system in $CHCl_3$ and compared them with those obtained for unsymmetric ADD–DAA/DDA–AAD and symmetric DAD–ADA H-bonded systems.

It is clear that the use of an ADD–DAA/DDA–AAD H-bonding pattern, providing moderate K_{ref} and high *EM* values, leads to cyclic tetramer assemblies that persist as the main species in solution over a much broader concentration range. Only at very low concentrations is $c\mathbf{M}_4$ dissociated as a monomer, but no other associated species is seen to compete within the 10^{-8}–$10^5\,M$ concentration range, thus 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 grow at the expense of the $c\mathbf{M}_4$ species. On the other hand, the DDD–AAA pattern 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 in the 10^{-5}–$1\,M$ range; however, 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, no associated species is seen below 10^{-4} M. At intermediate concentrations, a distribution of small oligomers, among which $c\mathbf{M}_4$ is one of the main species, is observed. In analogy to the DDD–AAA system, when the concentration is increased above 1 M, the higher-order linear oligomers dominate. However, high-molecular-weight distributions are attained much faster with the DDD–AAA system than with the DAD–ADA H-bonding pattern, as a result of a K_{ref} value that is about four orders of magnitude higher. It should be noted that the quantitative results derived from Figure 3b must be strictly applied to the specific monomer geometries investigated in this study.

In summary, in this study we have been able to dissect and analyze independently the contributions of the H-bonding strength between complementary nucleobases, as explained by the Jorgensen model, and the intrinsic chelate effect that they exert in cyclic systems. The results presented clearly demonstrate that cyclic assemblies constructed from symmetric DAD–ADA H-bonding pairs are much less stable than the homologues from unsymmetric ADD–DAA or DDA–AAD pairs. On one hand, the DAD–ADA bonding pattern reduces considerably the enthalpy of intermolecular association owing 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; in our case, the *EM* was reduced by about three orders of magnitude. Our conclusions could in principle be extended to many linear or cyclic supramolecular systems assembled through 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 EMs of the cyclization process. In other words, the binding interaction should also contribute to the preorganization of the system towards a specific structure, thus reducing the degrees of freedom of any other competitive supramolecular species. If, on the other hand, linear supramolecular polymers are pursued, a strong symmetric binding interaction would be the best choice to minimize chelate cooperativity and hence the tendency of the supramolecular system to form undesired cycles.

Acknowledgements

Funding from the European Union (ERC-Starting Grant 279548) and MINECO (CTQ2011-23659 and CTQ2014-57729-P) is gratefully acknowledged.

Keywords: chelate effect · effective molarity · noncovalent synthesis · nucleoside self-assembly · supramolecular chemistry

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 223–227
Angew. Chem. **2016**, *128*, 231–235

[1] a) L. J. Prins, D. N. Reinhoudt, P. Timmerman, *Angew. Chem. Int. Ed.* **2001**, *40*, 2382–2426; *Angew. Chem.* **2001**, *113*, 2446–2492; b) D. N. Reinhoudt, M. Crego-Calama, *Science* **2002**, *295*, 2403–2407.

[2] D. González-Rodríguez, A. P. H. J. Schenning, *Chem. Mater.* **2011**, *23*, 310–325.

[3] a) J. D. Badjicä, A. Nelson, S. J. Cantrill, W. B. Turnbull, J. F. Stoddart, *Acc. Chem. Res.* **2005**, *38*, 723–732; b) Focus Issue on Cooperativity, *Nat. Chem. Biol.* **2008**, *4*, 433–507; c) C. A. Hunter, H. L. Anderson, *Angew. Chem. Int. Ed.* **2009**, *48*, 7488–7499; *Angew. Chem.* **2009**, *121*, 7624–7636; d) G. Ercolani, L. Schiaffino, *Angew. Chem. Int. Ed.* **2011**, *50*, 1762–1768; *Angew. Chem.* **2011**, *123*, 1800–1807.

[4] 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.

[5] a) M. Fujita, M. Tominaga, A. Hori, B. Therrien, *Acc. Chem. Res.* **2005**, *38*, 371–380; b) M. J. Mayoral, N. Bilbao, D. González-Rodríguez, *ChemistryOpen* **2015**, DOI: 10.1002/open.201500171.

[6] a) G. Ercolani, *J. Phys. Chem. B* **1998**, *102*, 5699–5703; b) G. Ercolani, *J. Phys. Chem. B* **2003**, *107*, 5052–5057; c) G. Ercolani, *Struct. Bonding (Berlin)* **2006**, *121*, 167–215; d) H. Sun, C. A. Hunter, C. Navarro, S. Turega, *J. Am. Chem. Soc.* **2013**, *135*, 13129–13141.

[7] a) T. F. A. de Greef, E. W. Meijer, *Nature* **2008**, *453*, 171–173; b) B. A. Blight, C. A. Hunter, D. A. Leigh, H. McNab, P. I. T. Thomson, *Nat. Chem.* **2011**, *3*, 244–248; c) S. K. Yang, S. C. Zimmerman, *Isr. J. Chem.* **2013**, *53*, 511–520.

[8] a) S. Sivakova, S. J. Rowan, *Chem. Soc. Rev.* **2005**, *34*, 9–21; b) J. L. Sessler, C. M. Lawrence, J. Jayawickramarajah, *Chem. Soc. Rev.* **2007**, *36*, 314–325; c) M. Fathalla, C. M. Lawrence, N. Zhang, J. L. Sessler, J. Jayawickramarajah, *Chem. Soc. Rev.* **2009**, *38*, 1608–1620.

[9] W. L. Jorgensen, J. Pranata, *J. Am. Chem. Soc.* **1990**, *112*, 2008–2010.

[10] a) 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; *Angew. Chem.* **2015**, *127*, 6884–6888; b) S. Romero-Pérez, J. Camacho-García, C. Montoro-García, A. M. López-Pérez, A. Sanz, M. J. Mayoral, D. González-Rodríguez, *Org. Lett.* **2015**, *17*, 2664–2667.

[11] See the Supporting information for further details.

[12] 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.

[13] a) R. L. Ornstein, J. R. Fresco, *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 5171–5175; b) S. N. Rao, P. A. Kollman, *Biopolymers* **1986**, *25*, 267–280.

[14] 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.

[15] 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). As previous studies on supramolecular systems have also shown (see Ref. [6d]), these *EM* values do not seem to follow any clear relationship with respect to solvent polarity or H-bond strength.

Received: September 21, 2015
Published online: November 20, 2015

