Proc. Indian Acad. Sci. (Chem. Sci.), Vol. 106, No. 5, October 1994, pp. 923–935.

Hydrogen bonding control of molecular self-assembly

JOHN FREDERICKS, JI YANG, STEVEN J GEIB and ANDREW D HAMILTON*
Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260, USA

Abstract. In this short review we describe approaches to the design and construction of synthetic molecules that mimic the process of self organization that is at the heart of biological complexity. Multi-subunit enzymes, viruses, and higher order DNA structures are formed by the non-covalent association of many smaller components. This self-assembly is controlled by the nature, number and orientation of interacting groups on the surface of the subunits. The central problem lies in overcoming the unfavorable entropy of multi-subunit association by significant enthalpic contribution from the binding of complementary regions on the subunits. We will place particular emphasis on the design of synthetic molecules that use hydrogen bonding interactions to control the formation of aggregates of well-defined structure.

Keywords. Hydrogen bonding; self-organization; molecular recognition; nanotechnology.

1. Introduction

Molecular organization is at the center of all biological processes. The non-covalent interaction of two or more molecular subunits can lead to the formation of large, well-defined and functional molecular aggregates. In Nature this strategy of self-assembly has several key advantages (Lindsey 1991).

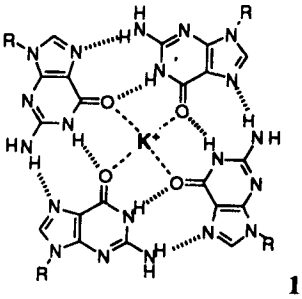
(a) The use of a few repeating subunits to create large structures reduces the amount of genetic information needed, compared to a single biopolymer chain.
(b) The weak aggregation of multiple substrates means that association and dissociation can be fast under different conditions leading to a greater control and potential switching of structure.
(c) The strict structural requirements imposed by the use of many small units on the final structure mean that any errors in subunits will be rejected as incompatible and so lead to an inbuilt error checking system.

The enormous complexity inherent and necessary in biological structures cannot be reached by individual biosynthetically derived units but requires the association of many individual components. The size, shape and makeup of the resulting aggregates will be controlled by the nature and orientation of the binding groups on the surface of the subunits. The arrangement of these binding regions represents chemical information that is imposed at the translation and folding stages but that

*For correspondence

is nonetheless sufficient to direct the non-covalent construction of the aggregate (Lindsey 1991; Whitesides *et al* 1992).

In nature, a good example of this self-organization process is seen in hemoglobin (Hendrickson 1977) where four subunits, two α- and two β-globin chains, self-assemble into a symmetrical tetramer that has oxygen binding and cooperativity properties very different from the individual monomers. Another example is provided by the photosynthetic reaction center from *Rhodopseudomonas viridis* which contains four different proteins that bring together four bacteriochlorophylls, two bacteriopheophytins and four hemes. The resulting supramolecular structure is an unsymmetrical multi-chromophore aggregate that spans the thylakoid membrane and possesses the function of one directional photoinduced electron transfer (Huber 1989). The principal interactions involved in these protein-protein associations are hydrophobic with hydrogen bonding and coulombic forces providing some additional binding energy and also conferring specificity (Janin *et al* 1990). A further advantage in the self-assembly strategy is that a switch or template element can be used to induce aggregation under a given set of environmental conditions. This is seen with insulin which can dimerize spontaneously but which forms a cyclic hexameric aggregate in the presence of zinc ions (Dodson *et al* 1993). In the DNA world, a similar example



1

is the guanine tetrad structure 1, found in teleomers, which self-organizes via eight hydrogen bonds around a template potassium ion (Sen and Gilbert 1990).

The presence of a cavity in these structures (the zinc binding site in insulin or the central carbonyl-flanked hole in 1) is essential for the switch mechanism to occur. In a similar way, an interior cavity with special binding or catalytic properties can result from controlled aggregation. The much-studied enzyme HIV protease (York *et al* 1993) is composed of two identical dimers that form a catalytic site at the protein interface (figure 1). A central cavity can also lead to interaction of the aggregate with a polymeric template that can control the make up and size of the assembly. For example, the viral coat proteins of the tobacco mosaic virus spontaneously self-organize (Stryer 1988) into a 34-unit, two-layered disc which in turn assembles into an elongated helical aggregate around a strand of template RNA (figure 2).

In this short review we will describe approaches taken by different groups, including our own, to the design and construction of synthetic molecules that self organize with some of the features seen in the biological structures outlined above. The use of small molecules mimics can lead to new insights into structural features of self assembly including how the size and number of an aggregate is controlled by the

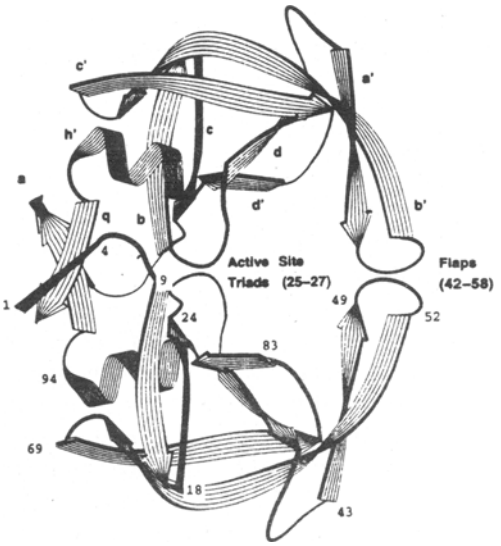


Figure 1. X-ray structure of HIV protease.

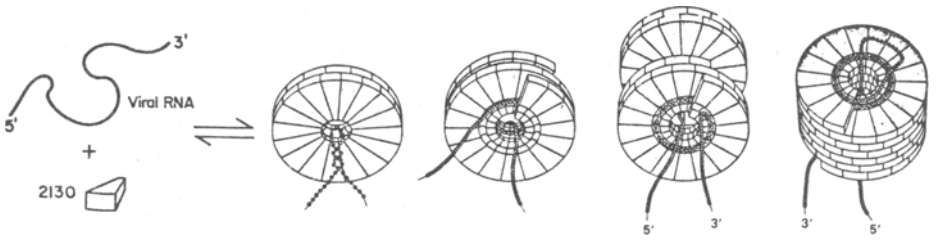


Figure 2. Self-organizing structure of tobacco mosaic virus (Stryer 1988).

position and orientation of the different interacting groups. A key question concerns the balance of enthalpic and entropic factors when several molecules form into a single assembly. This involves the loss of translational and rotational motion for each of the components and will result in a very large unfavorable entropy term. The size of this effect and the nature and number of binding interactions that must be deployed to overcome it are critical problems in the design of self-organizing molecules (Williams 1991).

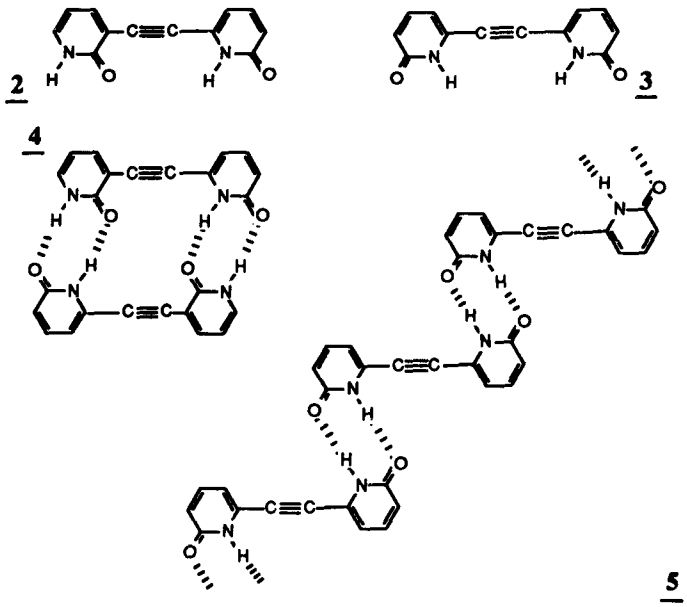
In his important review, Whitesides (Whitesides *et al* 1992) states that "nanostructures provide major unsolved problems in complexity and require new strategies and technologies for their synthesis and characterization. The solutions to these problems would be.... "essential elements in extending chemistry toward problems in materials science and biology." In addition to the insight gained into biological organization, self-assembling structures will find many applications in the development of artificial enzymes and in the construction of molecular electronic devices. The difficulty of using covalent synthesis to prepare large functional molecules suggests that non-covalent association can be an effective and programmable approach to collecting reactive groups at a single site. Some general approaches to these long term goals

have already been discussed (Hopfield *et al* 1989; Tecilla *et al* 1990; Whitesides *et al* 1992) in the construction of photoinduced electron and energy transfer systems.

2. Synthetic self-assembling structures

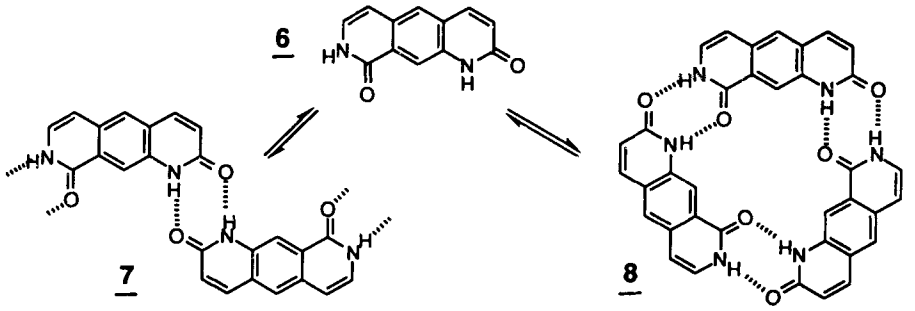
There has been a number of examples of designed synthetic molecules that self-organize into larger aggregates with defined structure and function (Lindsey 1991). In addition to our work on self-assembled multi-chromophore systems (Tecilla *et al* 1990), Sessler (Harriman *et al* 1991) and Aoyama (Aoyama *et al* 1991) have used hydrogen bonding to control the distances between porphyrins and electron or energy transfer acceptors in aggregates. These, however, represent functional aggregates formed by just two components. The number of studies on molecules that can potentially form aggregates of three or more subunits is much smaller.

The Wuest group have elegantly exploited the bidentate hydrogen bonding dimerization of 2-pyridones to form a family of structures that depending on the orientation of the interacting groups can form different types of aggregates (Persico and Wuest 1993). For example, two pyridone units were linked together with acetylene as a rigid spacer to form asymmetric and symmetric dipyridones 2 and 3, respectively. VPO, infrared spectroscopy, and X-ray crystallographic studies confirm that the asymmetric dipyridone 2 exists primarily as the dimer 4 in both solution and crystals. This self-association has been found to be particularly strong: the dipyridone exists as a dimer even at low concentrations ($3{\cdot}6 \times 10^{-4}$ M), and has a free energy of formation greater than 6·5 kcal mol^{-1} at 25°C. Crystalline symmetric dipyridone 3 is found to adopt the planar polymeric motif 5. The choice of solvent was found to affect the ability of these pyridones to self-associate. In methanol and other hydrogen-bonding polar solvents, 2 and 3 exist as solvated monomers, but they exist in self-associated



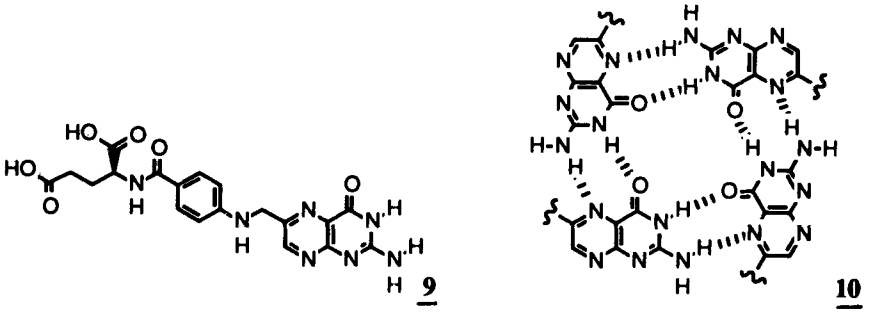
forms in chloroform (Gallant *et al* 1991a). To study the role of pre-organization in the formation of these aggregates, dipyridones, flexible analogs of 2 and 3 were synthesized and shown to form aggregates similar to 4 and 5 (Gallant *et al* 1991b). Attempts were also made to use dipyridone self-assembly as a template for self-replication via oxidative coupling of diyne analogs of 2 and 3. However, the strategy was not effective in orchestrating self-replication, perhaps because the dipyridones were too strongly self-associated in solution to bind to the subunits (Persico *et al* 1993).

The dipyridone hydrogen-bonding motif was also used by Zimmerman and Duerr (1992) in the design of a self-assembling cyclic trimer in solution. Pyridoquinoline 6 was synthesized because it had the potential to form either oligomeric aggregate 7 or a cyclic trimer 8.



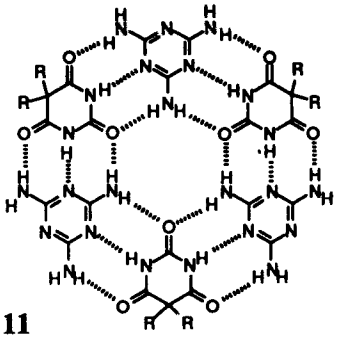
Using the Saunders–Hyne ^{1}H NMR method, they showed that 6 existed as the cyclic trimer in solution. The molecular weight of the aggregate determined by VPO in $CDCl_3$ at 35°C is 942 ± 80, within experimental error of the expected molecular weight of the trimer (MW = 1014). The trimer seems to exist even in polar solvents such as DMSO-d_6/$CDCl_3$. The trimer may be of greater stability than the oligomer because in the trimer two hydrogen bonds per molecule are formed, while the oligomer contains only $(2n + 2)/(n + 2)$ hydrogen bonds.

Folic acid salts also have been shown to self-assemble into cyclic aggregates. After observing that several deoxyguanidine derivatives form liquid crystal phases composed of cyclic tetramers which stack into chiral columnar aggregates (Mariani *et al* 1989; Bonazzi *et al* 1991). Spada and co-workers studied similar phenomena in folates. CD spectra and X-ray crystallographic data show that the dipotassium salt of folic acid 9 forms the cyclic tetramer 10 in water, which then stacks to form a chiral cylindrical

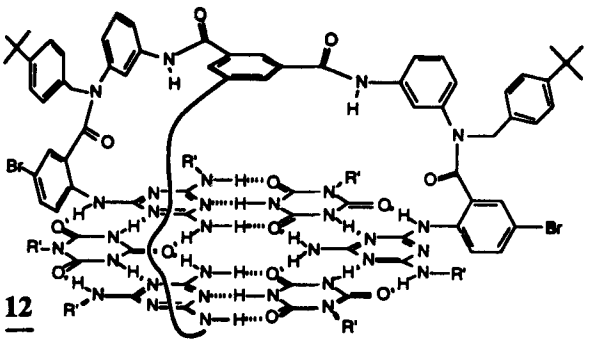


column. These columns then take up a parallel hexagonal order to form a liquid crystal mesophase.

Whitesides (Seto *et al* 1993; Seto and Whitesides 1993b) has elegantly exploited the melamine(M)-cyanuric acid (CA) lattice 11 to form a number of well defined aggregates, including the 3 + 1 complex shown in 12. Three melamine units were



11

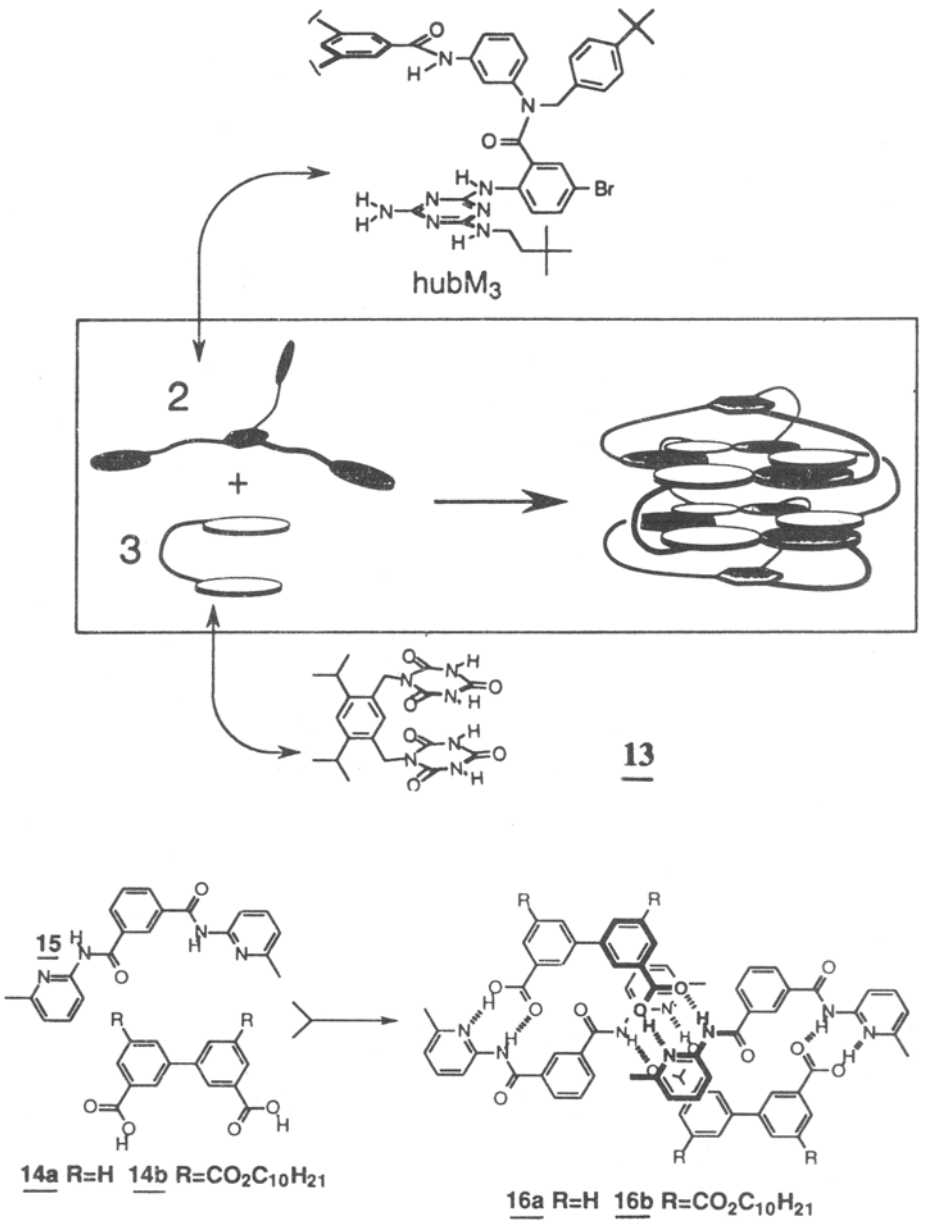


12

linked together by both rigid and flexible spacers to form the rigid $hubM_3$ and the flexible $flexM_3$. One equivalent of these units were each allowed to react with three equivalents of a cyanuric acid derivative (R'CA) to form the (1 + 3) complex $hubM_3(R'CA)_3$ 12 or $flexM_3(R'CA)_3$. Evidence for the formation of these complexes was obtained from solubility data, ^1H NMR, ^{13}C NMR, UV spectroscopy, GPC, and VPO. The molecular weight determined by VPO did not agree with the calculated weight expected for either complex. The molecular weight was 15–35% high for $hubM_3(R'CA)_3$ presumably due to some inter-complex hydrogen bonding. Despite the fact that these complexes can exist in four possible geometries, two C3 symmetrical and two unsymmetrical, the complex $hubM_3(R'CA)_3$ 12 is only found in the symmetrical form by ^1H NMR (Seto and Whitesides 1993a).

This hydrogen bonding motif has been extended to (2 + 3) complexes. In these aggregates, two units of $hubM_3$ were allowed to react with a subunit ($benzCA_2$) formed by linking two cyanuric acid groups through a benzene spacer. All evidence pointed to the existence of the (2 + 3) aggregates 13 in solution (Seto *et al* 1993).

Our own work has demonstrated in both solution and in the solid state, the formation of a cyclic 2 + 2 complex (Yang *et al* 1993) based on the strong interaction between a carboxylic acid and 2-acylaminopyridine. Using a conformationally restricted dicarboxylic acid (3,3'-biphenyldicarboxylic acid 14) and a *bis*-acylaminopyridine receptor 15 we have observed the preferential formation of a 2 + 2 complex 16. The X-ray structure of the complex (figure 3) shows a cyclic arrangement of alternating diacid-diamide units linked by eight hydrogen bonds (N...O distances, 2·70–2·93 Å). The overall shape is that of a figure-of-eight with one aminopyridine-carboxylic acid region stacking at a distance of ~ 3·5 Å to the corresponding region on the opposite side of the macrocycle. The integrity of the 2 + 2 aggregate is underlined by the absence of hydrogen bonds between adjacent aggregates. In $CDCl_3$ solution an equimolar mixture of 15 and 14b (each at 10 mM) at 25°C showed, in addition to the downfield shifts of the amide-H resonance due to hydrogen bonding,



upfield shifts of the pyridine-CH_3 and -5H resonances (0·5 and 0·3 ppm), compared to the uncomplexed subunits. These are consistent with a 2 + 2 aggregate in which one of the two pyridine rings in 15 is positioned above the benzoic acid subunit. Gel permeation chromatography in CH_2Cl_2 gave a peak with a shorter retention time (17·6 min; figure 4C) for an equimolar mixture of 15 and 14b (at 30 mM), compared to the individual components 15 (MW = 346) and 14b (MW = 611) which show sharp peaks with long retention times (21·0 and 20·4 min; figure 4, A and B). These results

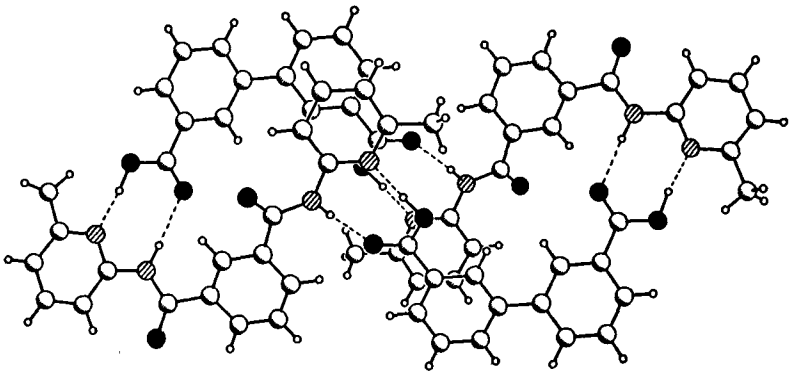


Figure 3. X-ray structure of 2 + 2 aggregate 16a.

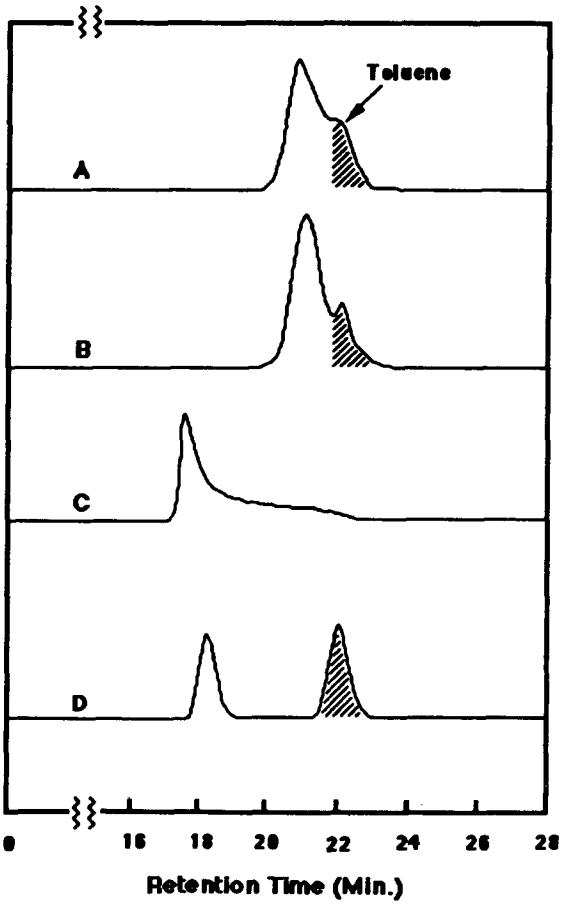
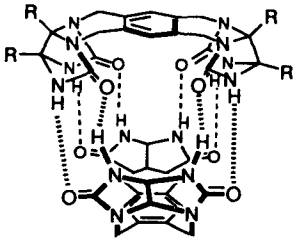


Figure 4. Gel permeation chromatograms with CH_2Cl_2 as eluent and toluene as standard for; A, 14b (3 mM); B, 15 (30 mM); C, 15:14b (1:1, 30 mM); D, standard (0·1 mM).

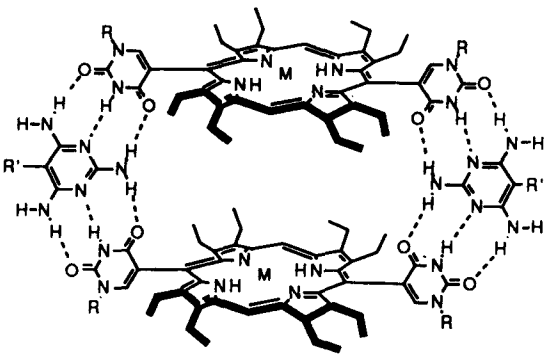
are consistent with the formation of a well-defined aggregate in solution with some dissociation of the aggregate on the column, as evidenced by the tailing visible in figure 4C. The molecular weight of the aggregate was estimated, by comparison to standards, (e.g. figure 4D, std. MW = 1704) to be 1970 ± 100, which is close to that calculated for the 2 + 2 aggregate 16b (MW = 1914).

3. Self-assembling structures with internal cavities

The above aggregates effectively show that, at least in simple systems, multiple subunits can be brought together via hydrogen bonding in a well-defined way. However, these are all relatively compact structures with no central cavity that might permit the binding of ions or the complexation of substrates. As discussed above, such features are a crucial part of self-organizing structures within biological systems and synthetic mimics must be developed to allow the imposition of switch or template functions. A number of examples has been reported where metal ions template the formation of helical complexes, however these depend exclusively on the metal as the determining factor with little interaction between the ligands (Koert *et al* 1990). In an elegant recent study, Rebek (Wyler *et al* 1993) has shown that carefully designed tetra-lactam derivatives can be induced to dimerize into a tennis ball-like structure, as in 17. Evidence based on NMR, VPO and X-ray crystallography confirm the formation of this unique complex. This represents a true self-organized cavity with the potential for inclusion of small substrates in the interior.



17



18

Lehn (Drain *et al* 1993) has also formed 2 + 2 aggregates based on porphyrin derivatives 18. The binding motif used was that between the 5-alkyluracil groups attached to the porphyrin ring, and two alkyltriaminopyrimidine (TAP) units. The aggregate was characterized using ^{1}H NMR, electrospray mass spectrometry, and VPO. Luminescence measurements also supported the structure. The uncomplexed porphyrin showed fluorescence quenching with increased concentration due to π-stacking. The self-assembled complex 18 did not show this quenching even up to concentrations of 5 mM. When 0·1% HCl was added the emission reverted to that

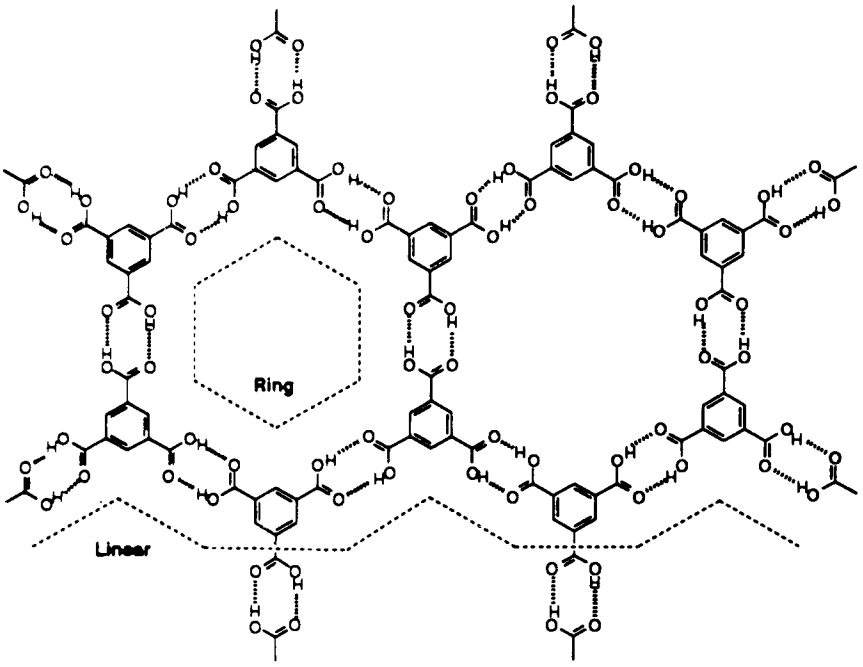


Figure 5. Extended sheet in X-ray structure of trimesic acid.

of the unassembled, protonated porphyrin. It was also found that the Zn complex binds 4,4'-bipyridine more strongly than Zn tetraphenyl porphyrin or Zn octethyl-porphyrin.

In the search for self-organizing molecules that possess a central cavity, some insight can come from consideration of the crystal structure packing patterns of hydrogen bonding molecules (Etter 1990). For example, the X-ray structure of trimesic acid shows a infinite sheet arrangement in which each molecule forms six hydrogen bonds to three neighboring residues (figure 5; Duchamp and Marsh 1969). The resulting hexagonal-lattice structure leads to large cavities each bounded by six trimesic acid molecules.

Isophthalic acid derivatives can form only partial elements of the sheet structure, such as the linear or ring motifs indicated in figure 5. For example, isophthalic acid itself forms the ribbon motif with a planar arrangement of phenyl and carboxylic acid groups (figure 6; Alcala and Martinez-Carrera 1972). The preference for formation of hydrogen bonded ribbon or sheet structures is well precedented in the solid state literature (Etter 1990). In the case of the isophthalic acid derivatives the ribbon motif allows both optimal formation of the bidentate hydrogen bonds and efficient packing interactions between the ribbons in vertical and horizontal directions. In order to direct formation of the cyclic motif 20 in the solid state it is necessary to disrupt the linear packing arrangement. This can be achieved by placing a bulky substituent in the 5-position of isophthalic acid, thus preventing easy alignment of the hydrogen bonded ribbons. The X-ray structure of 5-decyloxyisophthalic acid 19 (crystallized from THF/hexane; figure 7) shows the formation of a remarkable hexameric aggregate corresponding to the ring motif in figure 5 (Etter *et al* 1986). The six isophthalic acid

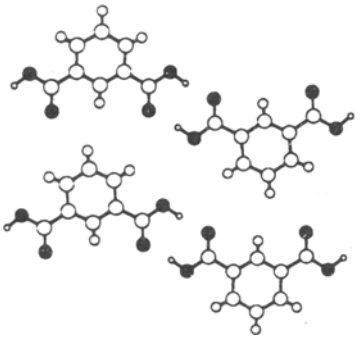


Figure 6. The X-ray of isophthalic acid.

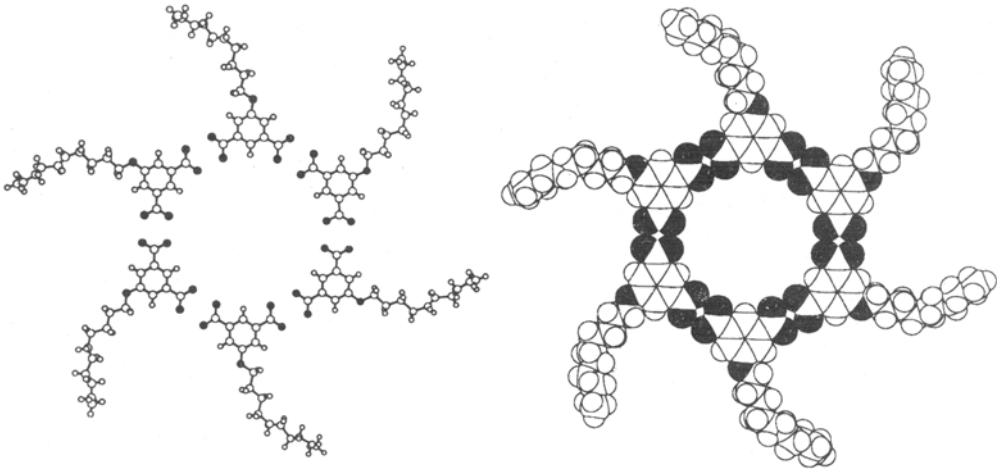
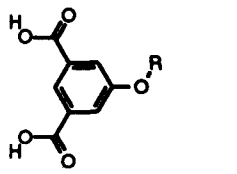


Figure 7. X-ray structure of 5-decyloxyisophthalic acid 19 (ball-and-stick and space filling).

molecules define a macrocyclic cavity that is 14 Å in diameter (from opposite isophthaloyl-2H sites) and is stabilized by 12 hydrogen bonds (O–O distance, 2·65 Å), as in 20.

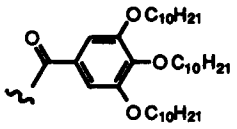
The isophthalate core of the structure is planar while the decyloxy chains maintain an alternating up-down arrangement.

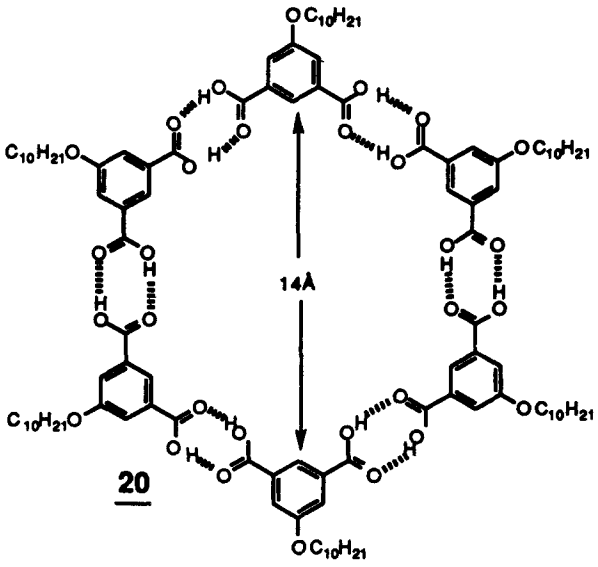
The question of whether these cyclic aggregates form in solution was investigated using the more soluble 5-substituted isophthalic acid derivative 21 by vapor phase osmometry, gel permeation chromatography and ^{1}H NMR. VPO measurements in toluene at 40°C gave molecular weights for the aggregate of 4600 (against a benzil standard) and 4900 (against a polystyrene standard) over a concentration range of 12–35 mM. The calculated MW for hexameric 21 is 4531. GPC using a toluene eluent showed a strong dependence on concentration. A major band with a short retention time was seen at high concentration (10 mM). At an intermediate concentration (2·5 mM) this band decreased and new bands with longer retention times appeared. Finally at a low concentration (0·1 mM) a new major peak with an even longer retention time was observed. These results are consistent with the preferential



19. R = $C_{10}H_{21}$

21. R =





20

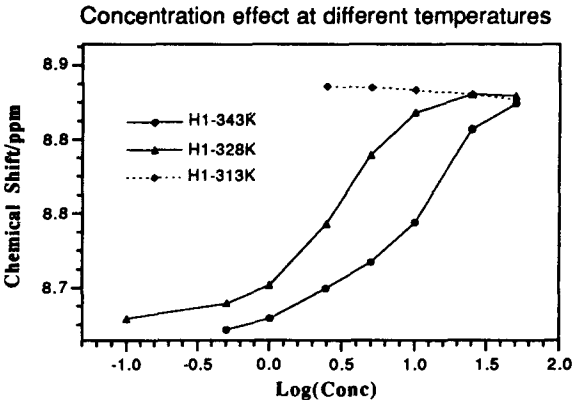


Figure 8. Plot of δ vs. log[21] at different temperatures.

formation of a hexameric aggregate in toluene at concentrations above 10 mM. The intermediate peaks presumably correspond to partial linear aggregates of 21 and the slowest moving band to fully dissociated 21.

The ^{1}H NMR spectrum of 21 in C_6D_6 (25 mM) at 55°C shows three sharp resonances at 8·89, 8·21 and 7·51 ppm from the isophthaloyl-2H and −4/6H and benzoyl-H, respectively. The position of the isophthaloyl-2H resonance changes with concentration, shifting upfield by 0·26 ppm over a 15–0·1 mM range (figure 8). Above 15 mM, however, there is little change, as expected for the formation of a discrete and stable aggregate at higher concentrations. The temperature dependence of this behavior is consistent with non covalent aggregation. At 70°C the dilution curve shows a shift to higher concentration reflecting destabilization of the aggregate. Whereas, at lower temperature (40°C) there is little change in chemical shift over a large concentration range consistent with a stable aggregate in solution.

References

Alcala R and Martinez-Carrera S 1972 *Acta Crystallogr.* **B28** 1671

Aoyama Y, Asakawa M, Matsui Y and Ogoshi H 1991 *J. Am. Chem. Soc.* **113** 6233

Bonazzi S, Capobianco M, De Morais M M, Garbesi A, Gottarelli G, Mariani P, Bossi M G P, Spada G P and Tondelli L 1991 *J. Am. Chem. Soc.* **113** 5809

Dodson E J, Dodson G G, Hubbard R E, Moody P C E, Turkenburg J, Whittingham J, Xiao B, Brange J, Kaarsholm N and Thogersen H 1993 *Proc. R. Soc.* **A345** 153

Drain C M, Fischer R, Nolen E G and Lehn J M 1993 *J. Chem. Soc., Chem. Commun.* 243

Duchamp D J and Marsh R E 1969 *Acta Crystallogr.* **B25** 5

Etter M C, Urbanczyk-Lipkowska Z, Jahn D A and Frye J S 1986 *J. Am. Chem. Soc.* **108** 5871

Etter M C 1990 *Acc. Chem. Res.* **23** 120

Gallant M, Viet M T P and Wuest J D 1991a *J. Am. Chem. Soc.* **113** 721

Gallant M, Viet M T P and Wuest J D 1991b *J. Org. Chem.* **56** 2284

Harriman A, Magda D and Sessler J L 1991 *J. Chem. Soc., Chem. Commun.* 345

Hendrickson W A 1977 *Trends Biochem. Sci.* **2** 108

Hopfield J J, Onuchic J N and Beretan D N 1989 *J. Phys. Chem.* **93** 6350

Huber R 1989 *Angew. Chem., Int. Ed. Engl.* **28** 848

Janin J and Chothia C 1990 *J. Biol. Chem.* **265** 16027

Koert U, Harding M M and Lehn J M 1990 *Nature (London)* **346** 339, and references therein

Lindsey J S 1991 *New J. Chem.* **15** 153

Mariani P, Mazabard C, Garbesi A, Spada G P and Tondelli L 1989 *J. Am. Chem. Soc.* **111** 6369

Persico F and Wuest J D 1993 *J. Org. Chem.* **58** 95

Sen D and Gilbert W 1990 *Nature (London)* **344** 410

Seto C T, Mathias J P and Whitesides G M 1993 *J. Am. Chem. Soc.* **115** 1321

Seto C T and Whitesides G M 1993a *J. Am. Chem. Soc.* **115** 905

Seto C T and Whitesides G M 1993b *J. Am. Chem. Soc.* **115** 1330

Stryer L 1988 In *Biochemistry* 3rd edn (New York: W H Freeman) p. 852

Tecilla P, Dixon R P, Slobodkin G, Alavi D S, Waldeck D H and Hamilton A D 1990 *J. Am. Chem. Soc.* **112** 9408

Whitesides G M, Mathias J P, Seto C T 1992 *Science* **254** 1312

Williams D H 1991 *Aldrichim. Acta* **24** 71

Wyler R, de Mendoza J and Rebek J Jr 1993 *Angew. Chem., Int. Ed. Engl.* **32** 1699

Yang J, Fan E, Geib S J and Hamilton A D 1993 *J. Am. Chem. Soc.* **115** 5314

York D M, Darden T A, Pedersen L G and Anderson M W 1993 *Biochemistry* **32** 1443

Zimmerman S C and Duerr B F 1992 *J. Org. Chem.* **57** 2215