

Heteroaromatic Modules for Self-Assembly Using Multiple Hydrogen Bonds

Steven C. Zimmerman, Perry S. Corbin

Department of Chemistry, University of Illinois, Urbana, IL 61801, USA

E-mail: sczimmer@uiuc.edu

Hydrogen bonding is a directional and moderately strong intermolecular force. Compounds that present multiple hydrogen-bond donor and acceptor groups have proven to be extremely important in creating new self-assembled structures. A review of several classes of organic compounds capable of multiple hydrogen-bond recognition is presented with a focus on the factors that contribute to complex stability.

Keywords: Hydrogen bonding, Molecular recognition, Self-assembly, Complexation, Heterocyclic compounds

| | | |
|----------|--|-----------|
| 1 | Introduction | 64 |
| 2 | Heteroaromatic Compounds with Complementary Hydrogen-Bonding Arrays | 64 |
| 2.1 | DAD and ADA Heteroaromatic Modules | 65 |
| 2.1.1 | Dimeric or Ditopic ADA and DAD Modules | 67 |
| 2.2 | DDA and AAD Heteroaromatic Modules | 69 |
| 2.2.1 | Dimeric or Ditopic AAD and DDA Modules | 71 |
| 2.3 | DDD and AAA Heteroaromatic Modules | 73 |
| 2.4 | ADAD Heteroaromatic Modules | 74 |
| 2.5 | DDAA Heteroaromatic Modules | 75 |
| 2.5.1 | Dimeric DDAA Heteroaromatic Modules | 76 |
| 2.6 | DAAD and ADDA Heteroaromatic Modules | 77 |
| 2.7 | AADDAA and DDAADD Heteroaromatic Modules | 78 |
| 3 | Factors Affecting Complex Stability and Specificity | 78 |
| 3.1 | Primary Hydrogen-Bond Strengths | 79 |
| 3.2 | Number of Hydrogen Bonds | 79 |
| 3.3 | Arrangement of Hydrogen-Bond Donor and Acceptor Groups | 80 |
| 3.4 | Preorganization | 83 |
| 3.5 | Protomeric Form: Effect on Complex Strength and Specificity | 85 |
| 3.6 | Other Factors | 86 |
| 4 | Selected Examples of Self-Assembling Systems | 87 |

| | | |
|---|--------------------------|----|
| 5 | Conclusions | 92 |
| 6 | References | 92 |

1

Introduction

Efforts to create chemical models of biochemical systems or processes have been termed biomimetic chemistry by Breslow [1]. Self-assembly, an essential process for creating and maintaining the organization of complex biological systems, has provided a special challenge for biomimetic chemists. This challenge is essentially one of chemical information storage and processing [2, 3]. For example, one can ask the question how can an organic molecule be constructed such that it will recognize itself or other molecules and assemble into a specific structure? One of the simplest approaches uses arrays of hydrogen-bond donor and acceptor groups and, thus, borrows directly from the genetic storage mechanism of DNA base-pairing.

This chapter, which serves as an updated and expanded version of a previous review [4], will focus on organic modules that recognize themselves or complementary modules by formation of multiple hydrogen bonds. We arbitrarily define multiple as more than two. This review is not intended to be comprehensive, but illustrative of basic concepts. As a result, many important studies are not covered, and a detailed description of the synthesis of these modules is beyond the scope of the discussion. However, the usefulness of these modules will clearly depend on the overall yield and number of steps needed to produce them, so synthetic accessibility is considered in the evaluation of several of the modules. Particular focus will be placed on the strength and specificity of complexation, with a detailed consideration of the factors that contribute to both. Several of the modules described herein were designed for use in host-guest systems or to answer fundamental questions concerning hydrogen-bond complexation. Many others were created explicitly for use in self-assembly. A few selected examples of self-assembling systems are described to illustrate the application of these heteroaromatic modules.

2

Heteroaromatic Compounds with Complementary Hydrogen-Bonding Arrays

As noted in the introduction, the use of heteroaromatic modules as recognition units for self-assembly is inspired by Nature's use of purines and pyrimidines as the storage units of genetic information. Numerous model studies of DNA and RNA base-pairing provide a considerable body of quantitative binding data (*vide infra*), as do the many host-guest studies that have targeted the nucleobases with heteroaromatic hosts [5–10]. Heteroaromatic compounds have several additional advantages, foremost of which is the geometrically well-defined, often linear array of hydrogen-bond donor (D) and

acceptor (A) groups presented on the edges of the heteroaromatic system. Furthermore, there are many general synthetic approaches to heteroaromatic compounds that can be used to prepare new modules. The main disadvantages of using heteroaromatic modules in self-assembly are that they are often poorly soluble and it is difficult to control and even establish the protomeric (tautomeric) form of many heteroaromatic compounds (see Sect. 3.5). In the sections that follow, examples of heteroaromatic modules and their corresponding complexes are presented.

2.1

DAD and ADA Heteroaromatic Modules

The DAD · ADA array is the most well-studied and most commonly used hydrogen-bonding motif. This particular array is not found in the common DNA base-pairs, but is present in base-pairs between 2-aminoadenine (A') and thymine (T) (e.g., 3 · 5, Fig. 1). Such A'T base-pairs are found naturally in the cyanophage S-2L [11] and, in synthetic oligonucleotides, have been found to increase duplex stability and specificity relative to adenine · thymine (AT) base-pairs [12, 13]. This property makes the A'T base-pair of interest in developing more effective antisense and antigene therapeutic agents. The increased stability of the A'T base-pair in oligonucleotides is reflected in a 0–2 and 2.5–6.5 °C increase in melting temperature (T_m) per A → A' substitution for duplex DNA and PNA · NA duplexes, respectively [12, 13]. This increase in T_m is consistent with model studies of base-pairing in organic solvents, in which the A · T base-pair has an association constant (K_{assoc}) of 90 M^{-1} in chloroform and the A'T base-pair (i.e., 3 · 5) has a K_{assoc} of 210 M^{-1} [5].

A series of DAD · ADA complexes and their corresponding K_{assoc} values in chloroform is displayed in Fig. 1. As can be seen, the values range from 65 to 900 M^{-1} . Although this is a considerable spread for complexes that contain the same number and arrangement of hydrogen bonds, the free energy (ΔG°) difference is only $1.5 \text{ kcal mol}^{-1}$. Furthermore, many of the association constants in Fig. 1 were measured in different laboratories. Thus, it is likely that some of the differences may be attributed to variability in the conditions for the binding studies. For example, complexes 6 · 7 and 20 · 21 are structurally very similar, yet their complexation free energies, measured in different laboratories, differ by ca. $1.0 \text{ kcal mol}^{-1}$. One variable that is usually not controlled in binding assays is the water content in the chloroform, which is likely to have a small effect on the K_{assoc} values, and a larger effect on the enthalpy of binding [20].

The most commonly used ADA modules contain the pyrimidine-2,4-dione nucleus. Not surprisingly, *N*-alkylation of thymine or uracil with an alkyl halide provides a simple, one-step method of functionalizing this module. Other ADA units include simple imides (e.g., 10), heterocycle 13, which was used by Kelly in a bisubstrate reaction template [17], and the anthryridinone or anthryridan units in 15 and 18, respectively. The latter modules are synthesized by double Friedländer condensation of 2,6-diaminopyridine-3,5-dicarboxal-

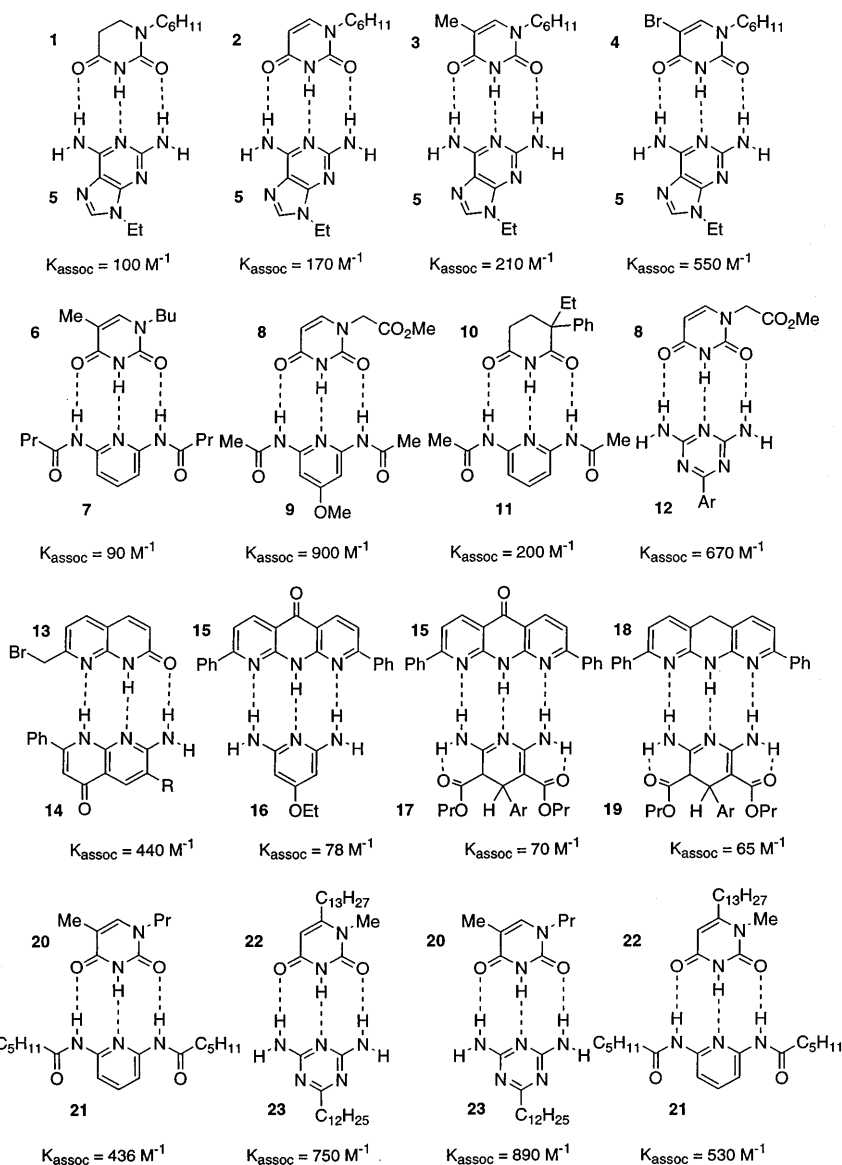


Fig. 1. Triply hydrogen-bonded complexes containing the ADA · DAD arrangement. Association constants are measured in chloroform-*d*. Complexes 5 · 1–4 [5]; 6 · 7 [14]; 8 · 9, 8 · 12 [15]; 10 · 11 [16]; 13 · 14 [17]; 15 · 16, 15 · 17, 18 · 19 [4, 18]; 20 · 21, 22 · 23, 20 · 23, 22 · 21 [19]

dehyde and acetyl aromatics or heteroaromatics, followed by either oxidative (15) or reductive workup (18) [21].

As noted above, 2-aminoadenine is a commonly used DAD module. In addition, the use of 2,6-diamidopyridines as the DAD complement to ADA

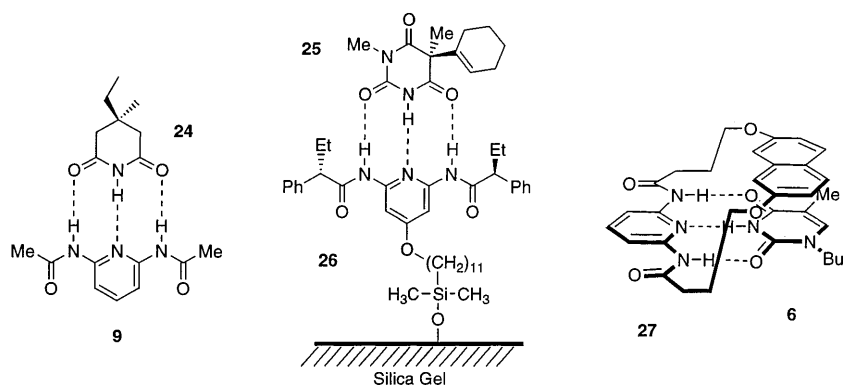


Fig. 2. Three ADA · DAD complexes featuring a 2,6-diamidopyridine unit as the DAD module

units has been described by Feibush, Karger et al. in the development of chemically bonded stationary phases for barbiturates, glutarimides, succinimides, and hydantoin [22]. The DAD · ADA pairing was clearly seen in an X-ray analysis of crystalline complex **9** · **24** (Fig. 2). Phase **26** was also effectively used for chiral separations, including the near base-line resolution of the enantiomers of hexobarbital (**25**). Hamilton has studied [23] thymine receptor **27**, which was shown to complex **6** in chloroform with $K_{\text{assoc}} = 290 \text{ M}^{-1}$ (vs 90 M^{-1} for **6** · **7**, Fig. 1). The added stability of **6** · **27** reflects π -stacking interactions in the complex. The naphthalene · thymine interaction was also seen in the solid-state structure. In these examples, the amide groups provide a convenient point of attachment to other compounds; however, unsymmetrically acylated diaminopyridines are not trivial to obtain. The 4-alkoxy group in compounds such as **9** provides another point of attachment, but its synthesis requires several steps. Related modules are the di- and triaminopyrimidines and di- and triaminotriazines, which are discussed in the next section.

2.1.1

Dimeric or Ditopic ADA and DAD Modules

There are two types of ditopic ADA and DAD modules, those synthesized by covalently linking modules, each containing one such hydrogen-bonding array, and those in which the hydrogen-bonding arrays are contained within a single heteroaromatic system. Along these lines, dimeric systems containing ADA subunits are readily available by alkylating suitable thymine or uracil analogs with an alkyl dibromide. Leonard's synthesis [24] of bis-thymine **28** (Fig. 3) illustrates this simple approach to dimeric modules. Interestingly, photolysis of **28** at 300 nm produces the *cis*-syn dimer **29** which projects a very rigid and somewhat splayed arrangement of the two ADA hydrogen-bonding arrays. In addition, Sessler has synthesized a rigid unit **30** as one in a series of "artificial dinucleotides", some of which form DNA-like duplexes [25]. These compounds are readily prepared by Sonagashira reactions with the appropri-

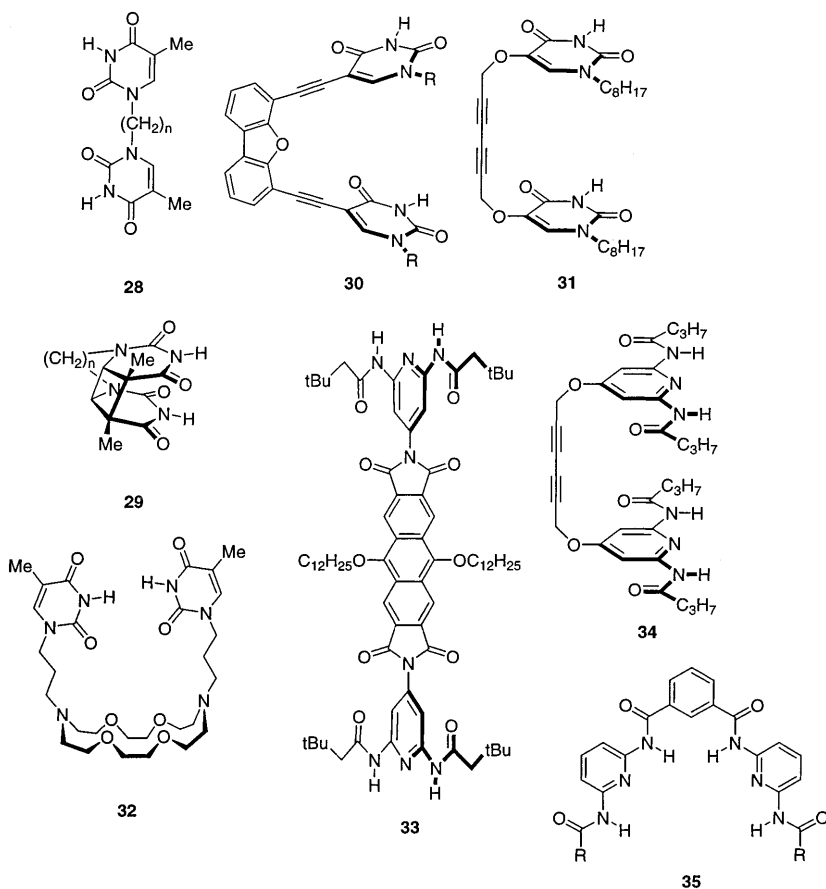


Fig. 3. Dimeric (ditopic) DAD and ADA modules constructed by covalently linking two identical modules

ate halogenated chromophores. Another example is Hamilton's linking of two 1-octyl-5-hydroxythymine units together with a diyne spacer to give **31** [26]. The connection of the subunits through the 5-alkoxy groups presumably allowed N-1 to carry the solubilizing octyl group. A final example is found in Gokel's extensive studies [27] of self-assembling ionophores (e.g., **32**), wherein base-pairing interactions create hosts capable of complexing complementary guest molecules.

Dimers of the DAD module are also illustrated in Fig. 3. Compound **34**, which holds two diamidopyridine units semi-rigidly with an interchromophore distance of about 10 Å, is analogous and complementary to **31** [26]. Likewise, Lehn developed the rigid, rod-like bis-diamidopyridine subunit **33** in which the hydrogen-bonding sites are directed along the long axis of the molecule [28]. As will be discussed in Sect. 5, such compounds are useful for creating polymeric

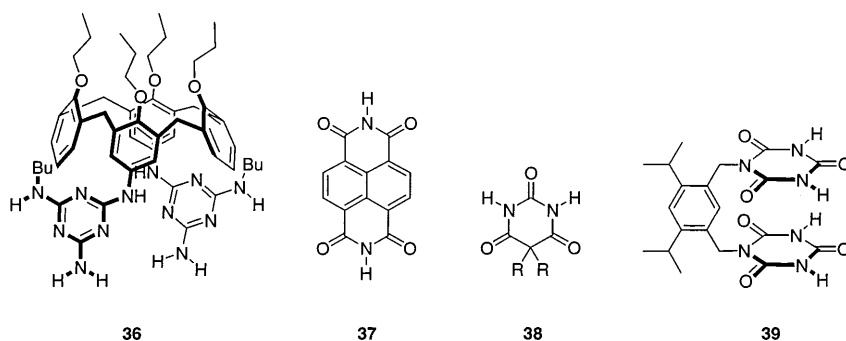


Fig. 4. Intrinsic ditopic modules containing two DAD or two ADA modules. Compounds **36** and **39** have two such modules linked by a semi-rigid scaffold

assemblies, whereas dimers such as **29–32**, **34**, and **35** are inclined to form closed, discrete aggregates with appropriate complementary partners.

In addition, the calix[4]arene-containing dimer of the triaminotriazine subunit, **36**, (Fig. 4) contains four recognition sites with two edges of each heteroaromatic group presenting the DAD hydrogen-bonding array [29]. Heterocycles such as the triaminotriazine unit, where the two hydrogen-bonding arrays are contained within a single heteroaromatic chromophore, may be termed *intrinsic* ditopic systems (Fig. 4). A ditopic ADA unit is found in commercially available naphthalene diimide (**37**) and in substituted barbituric acid (**38**), both of which have been used extensively as modules in self-assembly (*vide infra*). Compound **39** is analogous to **37** in presenting four recognition sites, but in this case, the ADA array is found in *N*-alkylated cyanuric acid derivatives [30]. The aminotriazene, barbituric acid, and cyanuric acid modules are readily available and easily derivatized, so the main effort is in synthesizing the dimers.

2.2

DDA and AAD Heteroaromatic Modules

The CG base-pair (i.e., **40** · **41**, R = ribose) is the prototypical DDA · AAD complex (Fig. 5). The AAD (cytosine) array is also found in 2-amidonaphthyridines (e.g., **42** and **44**). 2-Amino-5,7-dimethyl-1,8-naphthyridine is readily synthesized by the acid-catalyzed reaction of 2,6-diaminopyridine and acetylacetone [31] and is commercially available. Subsequent reaction with acetic anhydride readily affords **44**. In contrast, the DDA array is less common and less accessible synthetically. This motif is found primarily in the 6-amino-2(1*H*)-pyridone (**43**) [17], guanine (**41**, **47**), and 7-deazaguanine (**45**) [18] chromophores.

Guanine and guanosine are, of course, commercially available and can be derivatized in numerous ways. Along these lines, Sessler has used [32] a palladium-catalyzed Stille coupling to connect a porphyrin system to 8-bromoguanosine for use in a self-assembling, photosynthetic model system

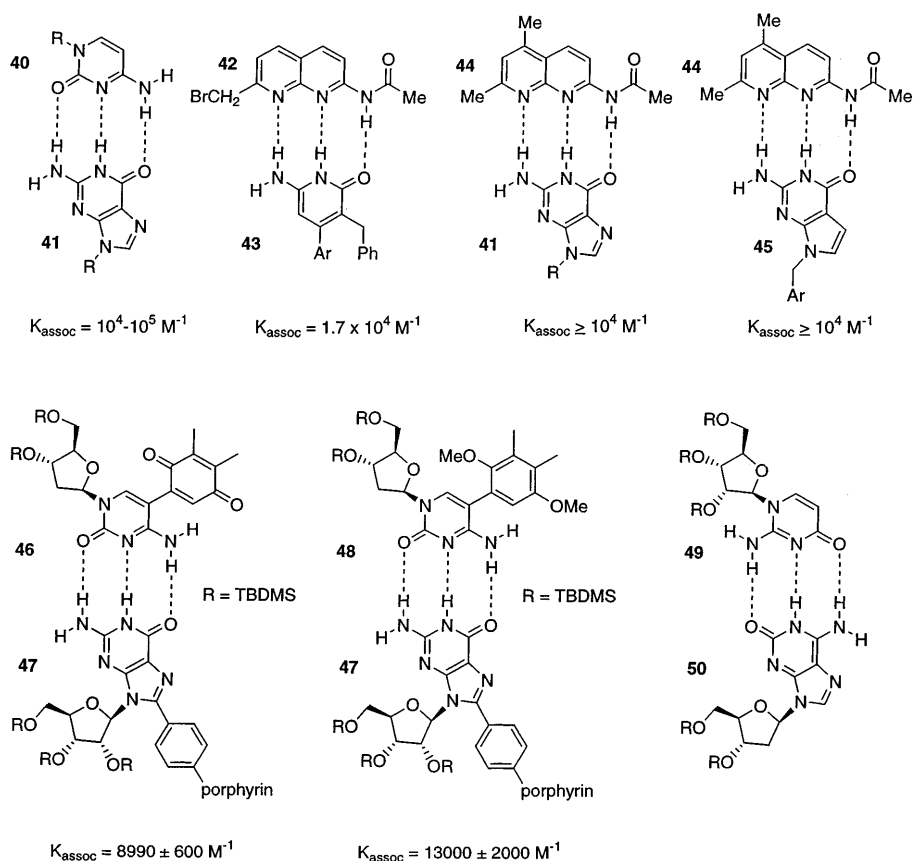


Fig. 5. Complexes containing the DDA · AAD motif. Association constants are measured in chloroform-*d*

(see 46 · 47 and 47 · 48). Likewise, isoguanosine can base-pair with isocytidine (see 49 · 50). An RNA · DNA 12-mer containing this non-natural base-pair exhibited the same melting temperature as the same sequence with a GC pair [33]. To our knowledge, the stability of isoG · isoC, isoG · C, or G · isoC complexes have not been measured in organic solvents.

The association constants listed in Fig. 5 appear to lie approximately in the $10^4 - 10^5 \text{ M}^{-1}$ range, which is two to three orders of magnitude higher than observed for the ADA · DAD systems. However, the difference may be even greater. Kyogoku and co-workers noted that their measured K_{assoc} values were only approximate due to the experimental method used [5], and Murray and Zimmerman found that 41 and 45 dimerized strongly in chloroform [4]. By extension, 43 and 47 are likely to dimerize, but this self-association was not taken into account when the K_{assoc} values in Fig. 5 were measured. Thus, the values reported are likely lower limits.

2.2.1

Dimeric or Ditopic AAD and DDA Modules

Several dimeric AAD and DDA modules have been reported that use porphyrin or a porphyrin analog as the scaffold (Fig. 6). For example, Sessler has covalently linked two guanine units to a sapphyrin (i.e., **51**) as a synthetic receptor for nucleotide phosphates [34] and has synthesized a bis-guanine analog of **47** [25]. Similarly, Hisatome has prepared [35] base-pair analogs by linking complementary nucleobases to aryl porphyrins (e.g., **52**). In the synthesis of **52** the guanine units are attached by an alkylation reaction as the last step. Evidence was presented for intramolecular base-pairing. In a related development, Sessler and Wang synthesized [25] and studied cytosine-guanine “heterodimer” **53** as part of their effort to create artificial duplexes (see **30**, *vide supra*). In this case, the bases are rigidly held apart so that base-pairing is forced to occur intermolecularly. Matile et al. have reported the synthesis of bis-guanine **54** as a first step toward a membrane spanning pore created by the tetramerization of polyphenylene-based guanine oligomers [36]. The authors noted solubility problems, which, in fact, are significant for most guanine-based compounds.

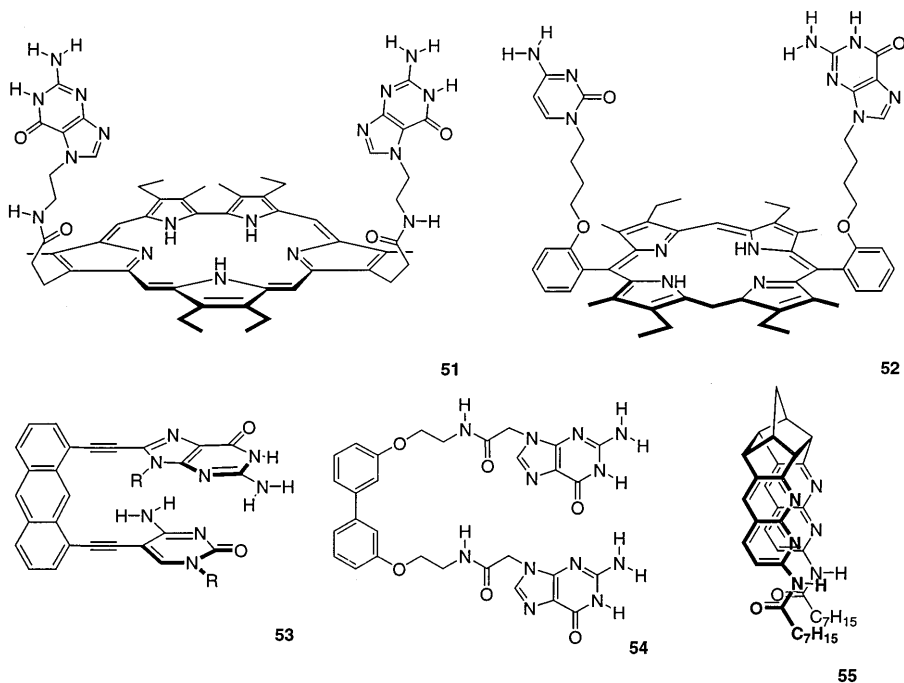


Fig. 6. Dimeric (ditopic) DDA and AAD modules constructed by covalently linking two identical modules

In contrast to the DDA units described above, there are comparatively few dimeric AAD modules. Zimmerman and co-workers have developed a convenient synthesis of 2,6-diaminopyridine-3-carboxaldehyde (DPC), which undergoes a Friedländer reaction with ketones to produce 2-amino-1,8-naphthyridines [37]. For example, the reaction of DPC with tetracyclo[6.3.0.0^{4,11}.0^{5,9}]undecane-2,7-dione followed by acylation with octanoic anhydride produced **55**. A number of diketones reacted similarly with DPC to produce bis-2-amino-1,8-naphthyridines with varying chromophore orientations.

In addition to the dimeric AAD/DDA units mentioned above, several intrinsic ditopic AAD/DDA modules have been reported (Fig. 7). In parallel, independent studies Lehn and Zimmerman developed tricyclic heteroaromatic compounds containing two DDA sites or two AAD sites [38–42]. Lehn has referred to these modules as Janus molecules [38]. Compounds **56** and **57** contain two cytosine nuclei (AAD array) fused to a central pyridine or benzene unit, respectively. Compound **56** is available in three steps from appropriate *N*-alkylureas [41], and **57** is available in approximately nine steps from resorcinol and 4-dodecylaniline [38]. The ditopic guanine analogs **58** and **59** contain two 5-amino-2(1*H*)-pyridone rings fused to a central pyrrole or benzene unit, respectively, and were designed as complements to **56** and **57**. Compound **58** is synthesized in five steps from methyl acetoacetate [41], and **59** is available in three steps from pyromellitic anhydride [38]. Whereas the 60° angle between the hydrogen-bonding arrays in the other heteroaromatic compounds in Fig. 7 is designed to produce closed-cyclic aggregates, the angle in **58** was intended to produce helical, polymeric aggregates [41].

Modules **60** and **61** are self-complementary, containing both a DDA and AAD hydrogen-bonding site. Both were designed to self-assemble into cyclic

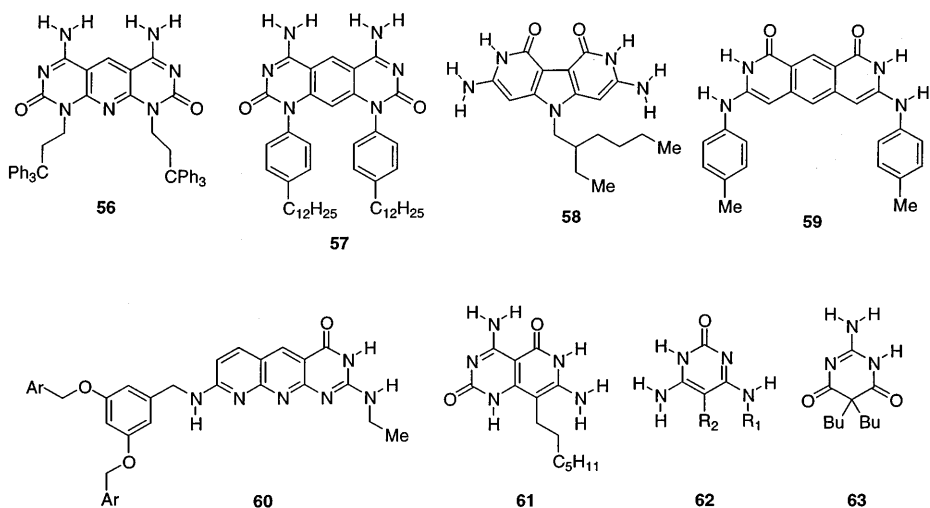


Fig. 7. Intrinsic ditopic modules containing DDA or AAD hydrogen-bond arrays

hexamers. Compound **60** is available from a multi-step synthesis [42]. Module **61** was prepared by Mascari in five steps starting from commercially available malononitrile dimer (2-amino-1-propene-1,1,3-tricarbonitrile) [43]. Janus modules **62** and **63**, both readily synthesized, contain a DDA and AAD hydrogen-bonding site [40]. Thus, they are self-complementary and complementary to one another. The self-assembly of most of these compounds has been studied and is discussed in Sect. 4.

2.3

DDD and AAA Heteroaromatic Modules

There are very few heteroaromatic modules that present the DDD or AAA hydrogen-bonding motif (Fig. 8). The latter is found in the 1,9,10-anthridine (1,8,9-triaza-anthracene) nucleus (e.g. **64**), which can be conveniently synthesized in a single step by a double Friedländer reaction of 2,6-diamino-3,5-pyridine dicarboxaldehyde and ketones [21]. One of the only DDD modules, dihydropyridine **65**, is also readily available via a single step Hantzsch synthesis using methyl carbamimidoyl acetate and 3-nitrobenzaldehyde [21]. The Hantzsch synthesis using 2-nitrobenzaldehyde affords the analogous dihydropyridine, which serves as a DAD module because it exists exclusively in the 3,4-dihydro form (i.e., **17**, Ar = 2-nitrophenyl) in both DMF and CDCl_3 [44]. Unfortunately, the dihydropyridines are prone to oxidation. In addition to this neutral DDD · AAA complex, Anslyn has demonstrated [45] that ethyl 2,6-diaminonicotinate in its protonated form (**66**), with a lipophilic tetraarylborate counterion, will dissolve in organic solvents and tightly complex **64** (Ar = phenyl). The authors note that although the measured pK_a values are consistent with complex **64** · **66**, the microenvironment of the complex could lead to a degree of proton transfer, which would lead to a complex with the DAD · ADA hydrogen-bonding arrangement. The cationic hydrogen bonding

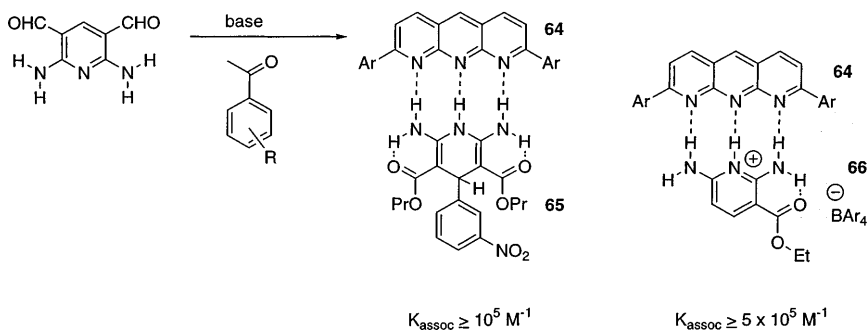


Fig. 8. Synthesis of the diaryl-1,9,10-anthridine (AAA) module **64** and its complex with DDD modules, dihydropyridine **65** and pyridinium **66**. Association constants are measured in chloroform-*d*

is expected to provide additional stability, and the oxidative instability of **65** is avoided by use of **66**.

2.4

ADAD Heteroaromatic Modules

One advantage of using modules with multiple hydrogen-bond donor and acceptor groups is the potential for creating self-assembled structures from multiple, complementary pairs of modules. Most self-assembly processes studied to date use a single recognition motif, but it is clear that more complicated assemblies will require a higher information content contained within two or more pairs of complementary recognition units. With two hydrogen-bonding sites, there are three modules possible that form two complexes: AA · DD and (AD)₂. As seen in the previous sections, a linear array of three sites can be arranged in six ways that form three hydrogen-bond motifs: ADA · DAD, AAD · DDA, and AAA · DDD. Four sites increase the number of arrays to ten, with six complexes possible, due to the fact that two of the modules are self-complementary: (ADAD)₂, (AADD)₂, ADDA · DAAD, ADDD · DAAA, DADD · ADAA, and AAAA · DDDD. The first three of these motifs have been studied and are discussed in this and the following section.

Meijer and co-workers have shown [46] that acylation of both diaminotriazines and diaminopyrimidines affords heteroaromatic systems containing the self-complementary DADA motif (see **67–70**, Fig. 9). Thus, well-established DAD modules can be readily converted into DADA units. The dimerization constants for **67–70**, measured in CDCl₃, vary considerably with the corresponding complexation free energies differing by over 4 kcal mol⁻¹ from strongest to weakest. This considerable variation for complexes with the same number of hydrogen bonds, and nominally the same hydrogen-bonding motif and geometry, will be discussed in more detail in Sect. 3. Two other DADA units have been reported as protomers of DDAA systems and, thus, will be discussed in the next section.

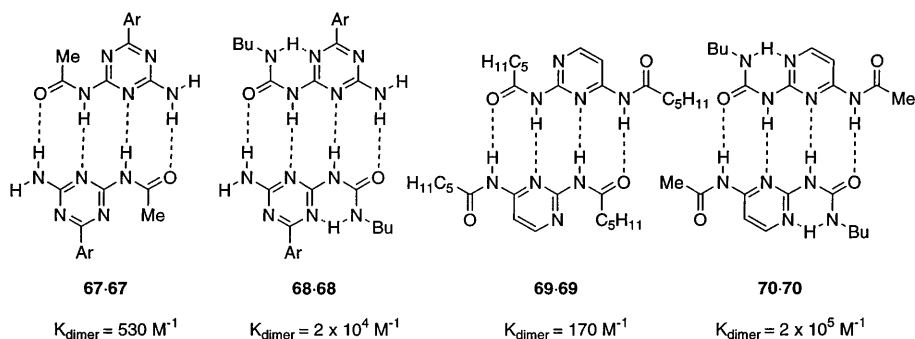


Fig. 9. Dimers of acylated diaminotriazine and -pyrimidine DADA modules. Association constants are measured in chloroform-*d*

2.5

DDAA Heteroaromatic Modules

Corbin and Zimmerman [47] and Meijer and co-workers [48, 49] have recently developed related DDAA modules (Figs. 10 and 11). The Corbin-Zimmerman module was created by efforts to meet several design criteria: (1) irrelevance of

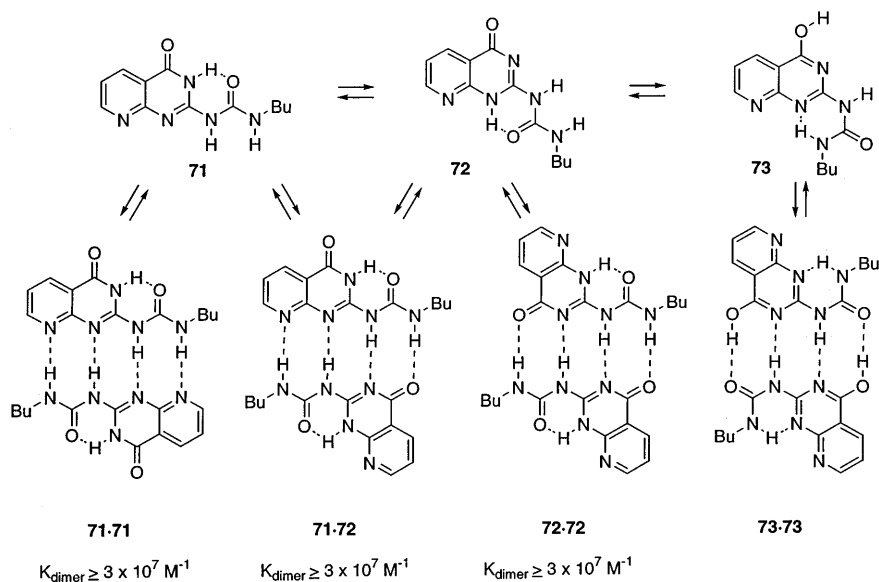


Fig. 10. Conformational, protomeric, and self-association equilibria of 71, a DDAA (ADAD) module

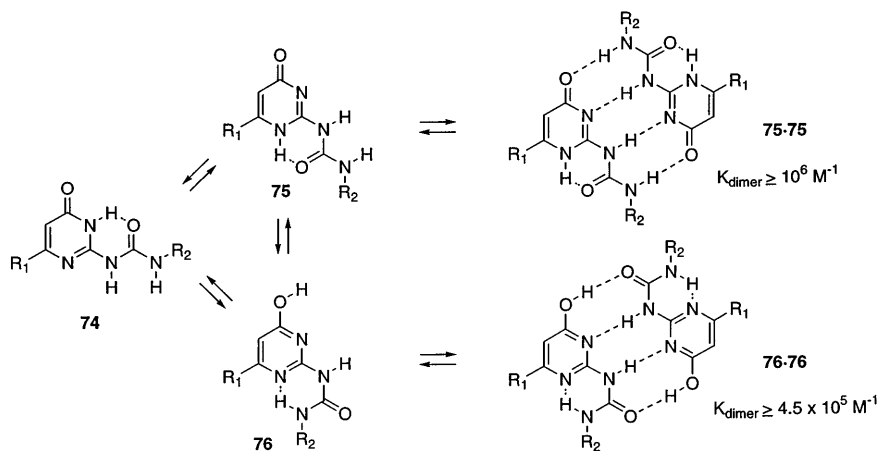


Fig. 11. Conformational, protomeric, and self-association equilibria of 74, a DDAA (ADAD) module

protomeric form, (2) very strong self-association, (3) inexpensive starting materials, (4) a scalable, short synthesis.

One of the problems with using heteroaromatic modules in self-assembly is controlling the protomeric equilibrium. The presence of undesired protomers may carry the wrong assembly instructions and will decrease the affinity with which the correct pairing occurs. This issue is taken up in greater detail in Sect. 3.

Compound **71** meets essentially all of the aforementioned design criteria (Fig. 10). Along these lines, compound **71** was synthesized in two steps from commercially available 2,6-diamino-3*H*-pyrimidin-4-one and 1,1,3,3-tetraethoxypropane [47, 50]. Protomer **71** was shown to exist in equilibrium with **72** in CDCl₃ and toluene-*d*₈, as ascertained by ¹H NMR studies. Moreover, both protomers associated forming homo dimers **71** · **71** (13%) and **72** · **72** (39%) and hetero dimer **71** · **72** (46%), with the relative percentages in toluene-*d*₈ shown in parentheses. The spatial arrangement of the alkyl substituent is maintained in each of these structures, which is important for its use in self-assembly. A minor protomer was tentatively assigned as **73**. This pyrimidino protomer contains the self-complementary DADA hydrogen-bonding motif and, thus, formed dimer **73** · **73** (2%).

Upon diluting a CDCl₃ solution of **71** to a concentration of 12 μM, no new peaks or shifting of existing peaks were observed in the ¹H NMR spectrum. Assuming ≥95% dimer formation at this concentration, a dimerization constant $K_{\text{dimer}} \geq 3 \times 10^7 \text{ M}^{-1}$ can be estimated. This is the highest stability constant measured to date for a neutral complex with four or fewer hydrogen bonds. As tightly as **71** and its protomers are associated, addition of a 2,7-diamido-1,8-naphthyridine led to complete dissociation of the dimers (see next section). The utility of **71** in creating discrete self-assembled structures in solution has been demonstrated [51].

Meijer and co-workers have extensively investigated module **74** [48]. A simple version of **74** (in which R₁ = CH₃ and R₂ = C₄H₉) is available in one step by reaction of methyl isocytosine and butyl isocyanate, respectively, whereas more soluble analogs (R₁ = C₁₃H₂₇; R₂ = C₄H₉ or Ar) are available in two steps from ethyl 3-oxohexadecanoate. Module **74** has several possible conformers/protomers. The three that were observed and characterized are 6(1*H*)-pyrimidinone **74**, 4(1*H*)-pyrimidinone **75**, and pyrimidinol **76** (Fig. 11). Of the three structures, **75** and **76** are capable of dimerizing. The X-ray analysis of **75** (R₁ = CH₃; R₂ = C₄H₉) shows dimer **75** · **75**. Interestingly, **76** (R₁ = C₆H₅; R₂ = C₄H₉) forms two types of crystals from chloroform, the first containing dimer **75** · **75** and the second dimer **76** · **76**.

2.5.1

Dimeric DDAA Heteroaromatic Modules

The ready availability of bis-isocyanates makes dimers of modules **71** and **75** straightforward to synthesize. Meijer and co-workers have pioneered this approach [49] to create hydrogen-bonded, supramolecular polymers of considerable size and molecular weight (see Sect. 4). A different dimeric

DDAA module developed by Sessler uses a single oxygen atom to accept two hydrogen bonds [52]. This rigid module (77, R = tris-*O*-TBDMS ribose), which is related to dimer 53, is shown in Fig. 12. Its self-assembled structure is shown here rather than in Sect. 4 to illustrate the non-standard pairing proposed to hold 77 · 77 together.

2.6

DAAD and ADDA Heteroaromatic Modules

Compound 71 can exist in additional conformations (e.g., 78, 79), as shown in Fig. 13. These conformations contain a non-self-complementary ADDA hydrogen-bond array, which is complementary to diamidonaphthyridine 80. In fact, two equivalents of 80 are sufficient to entirely convert the homodimers and heterodimer of 71 into complex 78 · 80 or 79 · 80 [47]. These complexes are, therefore, more stable than the already highly stable dimers. Preliminary computational work suggests that this is an intrinsic property of the complexes rather than an energetic preference for conformer 78 or 79. Also shown in Fig. 13 is the ADDA · DAAD complex formed between 80 and bis-2-pyridylurea 81 reported by Lüning [53]. Despite the structural similarity of

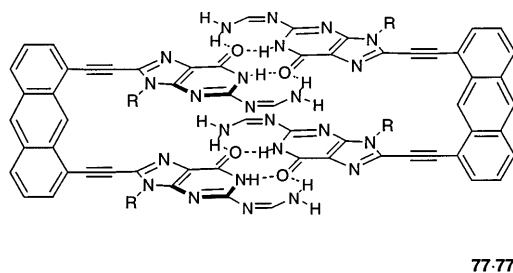


Fig. 12. Dimer formed from AADD module 77. Guanosine carbonyl acts as an AA unit by accepting two hydrogen bonds

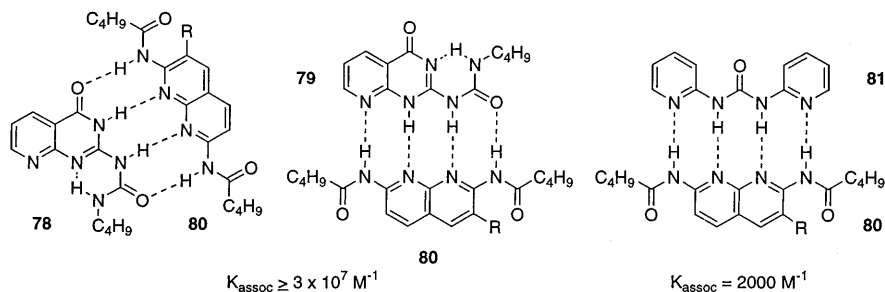


Fig. 13. Three complexes containing a DAAD · ADDA motif. In complexes with 78 (79) (R = H in 80) and in complex with 81 (R = CN in 80). Association constants are measured in chloroform-*d*

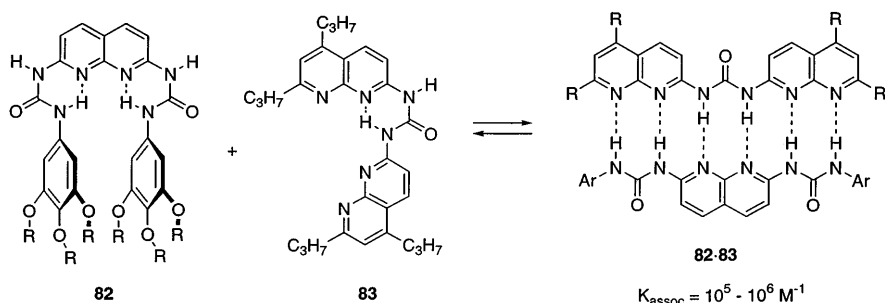


Fig. 14. Unfolding and association of DDAADD and AADDAA modules to form a complex with six hydrogen bonds. The association constant is measured in chloroform-*d*

$78 \cdot 80$ and $80 \cdot 81$, the K_{assoc} of $78 \cdot 80$ is only 2000 M^{-1} . This lower stability is due to a lower degree of preorganization, in particular an intramolecular hydrogen bond within **81**, which must be broken before complexation. Such issues are further addressed in Sect. 3.

2.7

AADDAA and DDAADD Heteroaromatic Modules

Corbin and Zimmerman have recently shown that a DDAADD module, bis-ureido-naphthyridine **82**, pairs with an AADDAA module, bis-naphthyridinyl-urea **83**, by forming complex $82 \cdot 83$, which contains six hydrogen bonds (Fig. 14) [54]. The K_{assoc} in CDCl_3 is estimated to be between 10^5 and 10^6 M^{-1} . Compound **82** exists in a folded state with two intramolecular hydrogen bonds both in solution and in the solid state, and **83** contains one intramolecular hydrogen bond in solution. Thus, the formation of six intermolecular hydrogen bonds in $82 \cdot 83$ comes at the price of three intramolecular hydrogen bonds in the components. Complex $82 \cdot 83$ can be regarded as a model of an oligomeric or even polymeric naphthyridinylurea that may form a folded structure or self-associate through formation of multiple hydrogen bonds.

3

Factors Affecting Complex Stability and Specificity

The stability of self-assembled structures depends on the strength with which the individual recognition units bind one another. The complexes and dimers described in the previous sections cover a broad stability range. Because there are so many factors that affect the strength of complexation, one might wonder whether it is possible to predict a priori the stability constant for a particular hydrogen-bonded complex. Certainly, advances in computational approaches are bringing closer the day when one will be able to calculate absolute free energies of complexation before even setting foot in the laboratory [55]. As will be discussed below, there are some general trends based on considerable experimental data, and possibly even empirical relationships, that allow at

least a qualitative prediction of stability for several classes of hydrogen-bonded complexes.

The specificity with which individual components assemble is a critical factor in determining whether the correct self-assembled structure is formed. Often very strong complexes are formed selectively, but this need not be the case. One advantage of using a single recognition unit is the lack of competing complex equilibria in the self-association process. With two complementary recognition units one must be concerned with self-association of the components competing with complex formation.

3.1

Primary Hydrogen-Bond Strengths

A central determinant of stability of any hydrogen-bonded complex will be the strength of the individual hydrogen bonds. The primary hydrogen-bond strengths depend, in turn, on the acidity of the donor site and the basicity of the acceptor site. For example, the $\log K_{\text{assoc}}$ values for complexes between phenol and nineteen substituted pyridines linearly correlated with the gas-phase proton affinity of the same pyridines [56]. Recognizing the importance of primary hydrogen-bond strengths is one thing, but it is quite another to actually determine these strengths in the more complicated multiply hydrogen-bonded complexes described herein. Indeed, it is not generally possible to factor out the relative contribution of individual contacts in complicated complexes such as base-pairs (see Fig. 1).

However, it is possible to order the stability of structurally similar complexes based on measured $\text{p}K_{\text{a}}$ values or expected relative acidities or basicities. For example, the increase in stability observed in the complexes between aminoadenine 5 and uracil derivatives 1–4 (Fig. 1) was attributed to the increase in acidity for the uracil imino proton across the series from 1 to 4 [5]. Likewise, the observation that the K_{assoc} of 8 · 9 is ten times larger than that of 6 · 7 can be attributed to the increased basicity of the methoxy-substituted pyridine. However, as noted previously [4], the K_{assoc} values of 6 · 7 and 8 · 9 were measured in different laboratories, so a direct comparison must be made with caution.

3.2

Number of Hydrogen Bonds

In an early analysis, a linear correlation was made between the complexation free energy (ΔG°) in chloroform and the number of hydrogen bonds in eight hydrogen-bonded complexes formed from amide or imide and amino- or amidopyridine components [57]. From the linear correlation it was concluded that each hydrogen bond contributed approximately $1.2 \text{ kcal mol}^{-1}$ to the free energy of complexation.

Even if similar primary hydrogen-bond strengths are found in these complexes, such a correlation is surprising because it seemingly requires a linear correlation between the complexation entropy and the number of

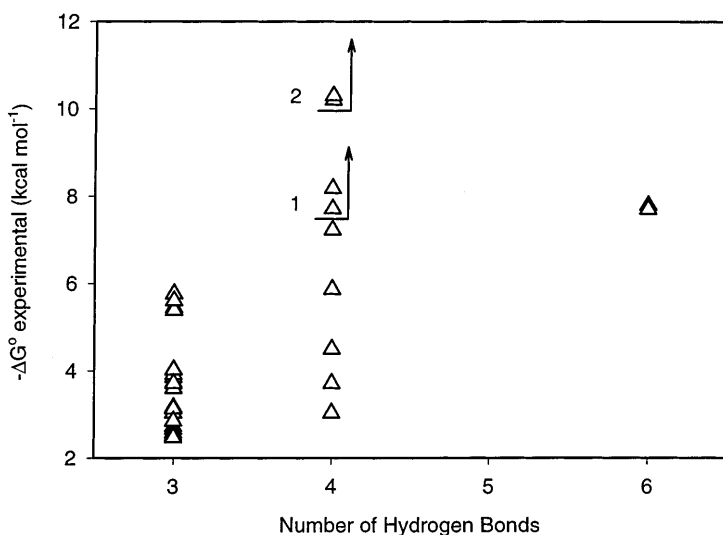


Fig. 15. Plot of free energies of complexation in chloroform-*d* vs number of hydrogen bonds for all complexes in this review. Data point 1 (75 · 75 and 76 · 76) and 2 (71 · 71 and 78 · 80) are plotted with ΔG° values that are lower limits. Actual values are likely to be considerably higher

hydrogen bonds. No attempts to generalize this type of correlation have been reported, and the analysis described above remains limited to carefully chosen complexes with similar types of hydrogen-bonding arrays. Indeed, there are now many examples where the more stable complex of a pair contains fewer hydrogen bonds. This situation can be seen graphically in Fig. 15, where the ΔG° values for each of the complexes discussed in this review are plotted against the number of hydrogen bonds.

3.3

Arrangement of Hydrogen-Bond Donor and Acceptor Groups

Jorgensen observed that base-pairs 2 · 5 and 6 · 7 have similar K_{assoc} values (ca. 100 M^{-1}), but that two other triply hydrogen-bonded complexes (40 · 41 and 42 · 43) displayed K_{assoc} values that were at least two orders of magnitude higher [58]. The differences, which were confirmed computationally, were attributed to secondary hydrogen bonding or secondary electrostatic interactions arising from the different arrangements of the hydrogen-bond donor and acceptor groups within the complexes (see Fig. 16). In short, Jorgensen proposed that the adjacent donor and acceptor groups are sufficiently close to make a significant contribution to complex stability. In the DAD · ADA motif of complexes 2 · 5 and 6 · 7, each of the secondary contacts are repulsive and destabilize the complex. In the DDA · AAD motif of complexes 40 · 41 and 42 · 43 two attractive secondary contacts offset two repulsive ones, which explains the higher stability of the DDA · AAD complexes. Jorgensen further

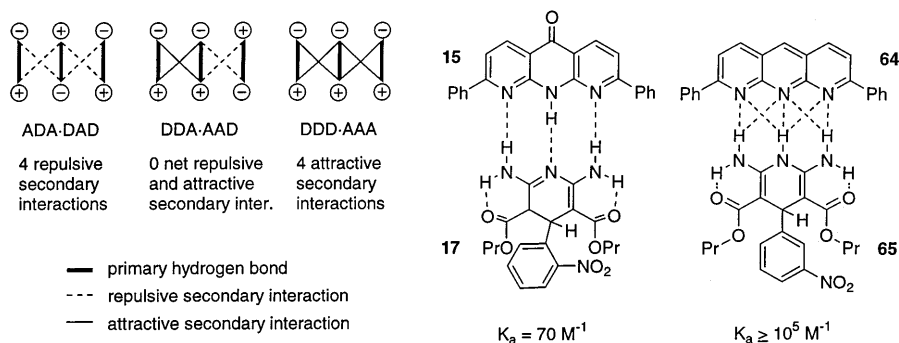


Fig. 16. Jorgensen secondary electrostatic model and its illustration with dramatically different K_{assoc} seen for DAD · ADA and DDD · AAA complexes

proposed that a complex with the AAA · DDD motif would be expected to be particularly robust as a result of four attractive secondary hydrogen bonds.

As described above, Murray and Zimmerman reported [18] the first example of an AAA · DDD complex (**64** · **65**). The ability to switch the protomeric form of **17** to **65** by a remote substituent effect, and construct ADA and AAA modules **15**, **18** and **64** with the 2,7-diphenylanthryridine nucleus, allowed for a careful comparison of structurally similar complexes. The $>10^3$ -fold difference in K_{assoc} for **15** · **17** (**18** · **19**) and **64** · **65**, and the high stability of the latter, was consistent with the Jorgensen proposal. The four attractive interactions in **64** · **65** are shown as hydrogen bonds in Fig. 16. The K_{assoc} trends in Figs. 1, 5, 8, 10 and 11 for closely related complexes are generally in agreement with the proposal that the arrangement of the hydrogen-bond donor and acceptor groups in a given module is a critical factor in determining complex stability. For example, the DDAA dimer **75** · **75** has a K_{assoc} value ca. 4 orders of magnitude higher than the related DADA dimer **76** · **76** in DMSO/chloroform mixtures, although their K_{assoc} are indistinguishably strong in pure chloroform [48]. Furthermore, systems with two hydrogen bonds and “overhanging” secondary interactions, which are not covered by this review, are also consistent with the proposal [58, 59]. On the other hand, there are some clear exceptions. For example, DAAD · ADDA complex **78** · **80** is strong enough to dissociate the AADD dimers of **71** even though the former contains two net repulsive secondary interactions, while the latter have two net attractive secondary contacts. Furthermore, the wide range of K_{assoc} values in Fig. 9 indicates that other factors are significant in determining complex stability because each of these complexes has the same number of hydrogen bonds and the same ADAD array.

The extension of the secondary hydrogen-bonding concept to peptides was investigated by Gellman [60], who tested the hypothesis that glycine dimer **84** would be less stable than complex **85** · **86** when constrained to be planar (Fig. 17). Foldamers **87** and **89**, cleverly designed to test this hypothesis, provided a scaffold wherein a single intramolecular hydrogen bond preorganizes the system for the second internal hydrogen bond, which is of the glycine

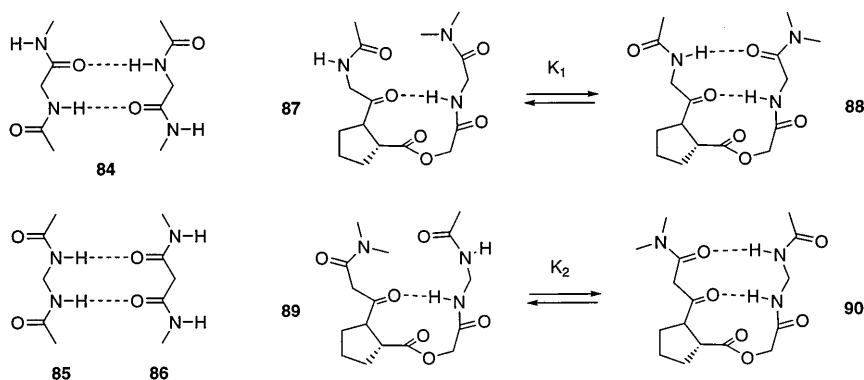


Fig. 17. A test of Jorgensen's secondary electrostatic hypothesis in flexible peptide models

dimer-type (**88**) or analogous to that in complex **85** · **86** (**90**). The experimentally determined K_1 and K_2 values (Fig. 17) were nearly identical, and the authors concluded that dipole–dipole repulsions within the arms of **90** favor alternative conformations. In relation to the heteroaromatic modules described herein, in the chemical synthesis of compounds such as **64** and **65**, the energy price is already paid to align the dipoles.

The tremendous scatter seen in the plot of ΔG° versus number of hydrogen bonds (Fig. 15) is clearly due in part to the differences in secondary hydrogen bonding. This raises the question of whether consideration of both factors could lead to the ability to predict complexation free energies. Along these lines, Schneider has proposed [61], based on multiple linear regression analysis, that there is a linear relationship between the observed and calculated free energies of complexation if each primary hydrogen bond contributes $1.9 \text{ kcal mol}^{-1}$ to the total energy and each secondary contact $\pm 0.69 \text{ kcal mol}^{-1}$, with the sign of the latter depending on whether it is attractive or repulsive. The author's plot of $\Delta G_{\text{calcd}}^\circ$ versus $\Delta G_{\text{expt}}^\circ$ for 58 different complexes gave a slope of 0.84, with an $R^2 = 0.834$, and an average difference between ΔG° values of just $\pm 0.4 \text{ kcal mol}^{-1}$. This remarkable agreement suggests that the stability of multiply hydrogen-bonded complexes can be predicted a priori using these empirical increments. However, complexes have been reported where the $\Delta G_{\text{calcd}}^\circ$ and the $\Delta G_{\text{expt}}^\circ$ values are not in good agreement [46, 53].

A plot of $\Delta G_{\text{calcd}}^\circ$ versus $\Delta G_{\text{expt}}^\circ$ is shown in Fig. 18 for the complexes described in this review, along with the best-fit line (solid) and the line expected for an ideal one-to-one correlation (dotted). Although a few points fall on the best-fit line, the slope is 0.68 and the $R^2 = 0.535$. Furthermore, the four DDA · AAD complexes that have ΔG° values that fall on the solid line (data points 1 in Fig. 18) are likely, as described in Sect. 2.2, to have complexation energies in excess of the reported values. This is represented by the up-arrow in Fig. 18. Data point 2 for complex **78** · **80** is plotted with $K_{\text{assoc}} = 3 \times 10^7 \text{ M}^{-1}$. Because this is a lower limit, its true K_{assoc} exceeds its calculated K_{assoc} of $3.6 \times 10^4 \text{ M}^{-1}$ by several orders of magnitude. The data in Fig. 18 suggest that

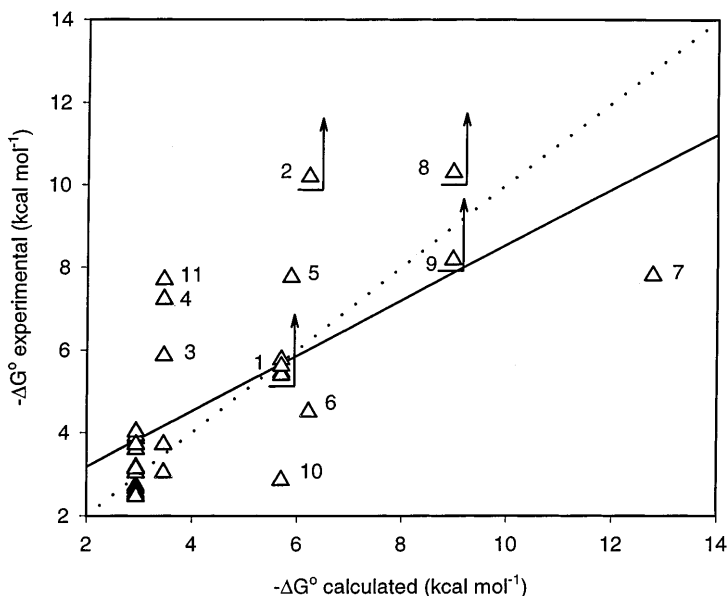


Fig. 18. Plot of free energies of complexation for all complexes in this review vs calculated ΔG° values using [61]. The *solid line* represents the linear regression fit of the data and the *dotted line* is expected for a 1:1 correlation of experimental and calculated values. Data points: 1 (DDA · ADD complexes), 2 (78 · 80), 3 (68 · 68), 4 (70 · 70), 5 (35 · 38), 6 (80 · 81), 7 (82 · 83), 8 (71 · 71), 9 (75 · 75), 10 (41 · 91), 11 (76 · 76). The *up-arrows* indicate points whose ΔG° values are lower limits

the previous analysis is not general and that other factors beyond the number of primary and secondary hydrogen bonds must be considered to predict complex strength. In the case of 78 · 80, preliminary computational work suggests that the primary hydrogen bonding is especially strong [47]. Other factors that must be considered are discussed in the sections below.

3.4 Preorganization

As alluded to above, one of the central tenets of host-guest chemistry is the idea that the most effective hosts are rigidly held in the conformation needed for binding and are poorly solvated. Cram has stated this idea most explicitly in his principle of preorganization [62]. This principle states that complex stability will be highest when the host and guest are organized for binding and low solvation. It has been known for many years that crown ethers complex metal ions more strongly than their acyclic analogs (podands) [63]. Zimmerman and co-workers have quantified the free energy benefit of freezing single bond rotations in a host-guest (molecular tweezers) system [64]. Each free rotation cost ca. 1 kcal mol⁻¹ in complexation free energy. This value is comparable to the energy cost of freezing single bond rotations in intramo-

lecular reactions determined many years ago by Page and Jencks [65] and a more recent estimate of the entropic cost for complexation by Williams [66]. Clearly, systems that require many single bonds to be fixed will pay a serious energy cost. It should be noted that a very recent analysis of acyclic host-guest systems reached the opposite conclusion, suggesting a much lower complexation energy cost of $0.3 \text{ kcal mol}^{-1}$ per single bond rotation [67].

Several related complexes in this review differ in their degree of preorganization. For example, Meijer has attributed the higher stability of dimers $68 \cdot 68$ and $70 \cdot 70$ relative to $67 \cdot 67$ and $69 \cdot 69$ to the intramolecular hydrogen bond in the former complexes, which preorganizes the four hydrogen-bond donor and acceptor groups. Each intramolecular hydrogen bond increases the complex stability by ca. $1\text{--}2 \text{ kcal mol}^{-1}$. This value contains more than just the entropic advantage, as the non-hydrogen-bonded amides in the monomer likely adopt the enthalpically more stable *trans* form.

Complexes $80 \cdot 81$ and $82 \cdot 83$ are significantly weaker than expected; particularly the latter complex, which contains six hydrogen bonds and a DDAADD · AADDAA motif. An explanation for the weaker binding is simple. These components are poorly organized for complexation as a result of an intramolecular hydrogen bond in pyridyl urea **81** and naphthyridinyl urea **83**. In fact, the solid-state and solution structure of **82** reveal two additional intramolecular hydrogen bonds that must be broken prior to formation of $82 \cdot 83$. Thus, three hydrogen bonds must be broken to form the six hydrogen bonds in this complex.

In addition to intramolecular hydrogen bonding, other conformational issues may affect the level of preorganization. Along these lines, Hamilton found [68] that ethoxynaphthyridine **91** bound triacetyl guanosine **41** with $K_{\text{assoc}} = 126 \text{ M}^{-1}$ (Fig. 19), which is at least two orders of magnitude lower than the K_{assoc} of a very similar complex, $41 \cdot 44$. Murray and Zimmerman proposed [69] that when the ethoxy group is in its preferred orientation there are steric interactions with the guanine amino group (see $41 \cdot 91'$). Thus, there is an energy cost for producing the less stable conformation of the ethoxy group in

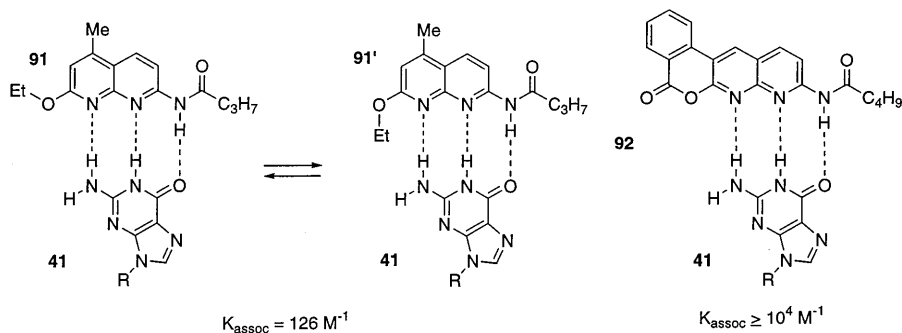


Fig. 19. Conformational equilibrium of ethoxynaphthyridine **91** and its complex with **41**. Compound **92** contains an oxy substituent but is constrained in a ring. Association constants are measured in chloroform-*d*

41 · 91. Evidence for this hypothesis came from studies of 92, which contains an alkoxy group in the 8-position that is “tied back” in a lactone ring. In contrast to 41 · 91, the K_{assoc} of 41 · 92 was $\geq 10^4 \text{ M}^{-1}$, which was expected of the DDA · AAD-type complex.

3.5

Protomeric Form: Effect on Complex Strength and Specificity

Watson and Crick suggested in 1953 that base substitution errors in DNA might occur as a result of the mis-pairing of minor protomeric forms of the four natural bases [70]. For example, the enol form of thymine (T_{enol}) is complementary to guanosine (G). Although such protomeric forms have been observed directly in the gas phase using model compounds [71], and in the X-ray structure of non-natural base-pairs [72], direct evidence for their importance in naturally occurring DNA is lacking. Nonetheless, many agree that the mis-insertion rate due to protomerism is likely to be in the range of 10^{-4} – 10^{-6} [72, 73]. How is this frequency of mis-insertion consistent with the high fidelity with which DNA is replicated (i.e., as few mistakes as one in 10^9 bases)? The seemingly inconsistent result is easily explained: DNA polymerases have a built-in proofreading function, and a second level of proofreading takes place post-replication.

Biomimetic chemists do not have the luxury of adding enzymes to check for mistakes in self-assembly. When modules contain the incorrect protomeric form, they carry misinformation that may translate into the wrong structure being assembled. Alternatively, a price may be paid in complex or assembly stability to shift the protomeric equilibrium toward the desired form. For example, Meyer has studied [44] the substituent effects on the protomeric equilibrium of 65 by ^1H NMR spectroscopy and has found that in dimethylformamide- d_7 (DMF- d_7) the 1,4-dihydro protomeric form exists exclusively. In chloroform- d (CDCl_3), however, a ca. 70:30 mixture of 1,4-dihydro (65) and 3,4-dihydro (17) protomers (slow exchange) were present. Thus, although complex 64 · 65 forms quantitatively from a 1:1 mixture of its components, an energy price must be paid to shift the protomeric equilibrium. Clearly, in this case, the energy price is very small ($<1 \text{ kcal mol}^{-1}$). In cases where the desired protomer is disfavored, the energy price will be higher.

Remarkably, few of the heteroaromatic compounds used as modules for self-assembly have had their protomeric equilibria examined. Many investigators simply assume the form they desire is the favored one. In part this is because determining the protomeric form with certainty is difficult and usually requires the synthesis of *N*- and *O*-methyl derivatives as bond-fixed analogs of the possible protomers. Even then spectroscopic analysis may be insufficient. Beak has shown [74] that many of the protomeric forms assigned to heteroaromatic compounds are incorrect because they were determined under conditions where the compounds were fully self-associated, which affects the protomeric equilibria.

Meijer and co-workers have studied [48] the protomeric equilibria of dimers 75 · 75 and 76 · 76, and found the equilibrium to be affected by the

electronic nature of the substituent (Fig. 11). For example, **76** was the exclusive form when $R_1 = CF_3$; $R_2 = C_4H_9$. In studies carried out in THF- d_8 , **76** was the preferred protomer in the monomeric state when $R_2 = C_4H_9$. In mixtures of $CDCl_3$ and DMSO- d_6 , the 6(1*H*)-pyrimidinone form (i.e., **74**) is preferred in the monomeric state. The dimerization constants for **75** · **75** and **76** · **76** were estimated to be very strong in $CDCl_3$, at least 4.5×10^5 and $10^6 M^{-1}$, respectively. In $CDCl_3$ /DMSO- d_6 mixtures, the dimerization constant for the DADA motif of **76** · **76** is ca. 10^4 -fold weaker than that for the DDAA motif in **75** · **75**. All these data suggest that the very strong dimerization of **75** is weakened by protomerism. However, the effect has not been quantified and is likely irrelevant for most applications given the very high stability of the dimer.

As described above, heteroaromatic module **71** was specifically designed [47] so that each major protomeric form would be capable of strong dimerization. To our knowledge, it is the first module designed to meet this criterion. As shown in Fig. 10, both the 1*H*- and 3*H*-protomers are indeed present in apolar organic solvents in homodimers **71** · **71** and **72** · **72** and heterodimer **71** · **72**. The K_{dimer} values in chloroform-*d* are immeasurably high, at least $3 \times 10^7 M^{-1}$. Although not by design, another heteroaromatic compound is known wherein two protomeric forms are capable of dimerizing. Specifically, 2(1*H*)-pyridone and 2-hydroxypyridine can each dimerize or form a complex together. However, the 2-pyridone dimerization constant is only ca. $100 M^{-1}$, and the heterodimer forms in a head-to-head arrangement whereas the homodimers are head-to-tail [47]. As described above, the dimers of **71** are expected to maintain a very similar spatial arrangement of the *N*-alkyl groups.

3.6

Other Factors

There are many other factors that will determine the strength with which heteroaromatic modules pair. Electrostatic effects, those beyond the primary and secondary hydrogen-bonding contacts discussed above, are likely to play an important role. Ion-pair hydrogen bonds are found to be stronger than neutral ones both in biological systems [75] and in abiotic complexes [76]. Jorgensen has noted that the CG base-pair may enjoy additional stabilization from the large dipole moment of the component bases, which interact favorably in the complex [58]. There is a possibility that complexes with favorable secondary hydrogen bonding will also have significant molecular dipolar attractions. For example, complexes **64** · **65** and **64** · **66** contain components whose molecular structures suggest large dipole moments that align in a favorable way in the complexes.

Another obvious factor that contributes to the pairing strength is the extent of solvation of the components and the complex itself. Chloroform is used in many complexation and self-assembly studies because it is not highly competitive for hydrogen-bond donor and acceptor sites, yet it is often one of the best solvents for dissolving heteroaromatic compounds. It is widely known that pairing strengths increase in even less polar solvents such as the

hydrocarbons and are greatly diminished in protic and other highly polar solvents such as dimethyl sulfoxide. Most designers of hydrogen-bonding modules rarely consider the issue of solvation, yet it may play a decisive role in determining the stability of an assembly. Hard data is hard to come by, but one might imagine that the modules described above containing the DDD and AAA hydrogen-bonding arrays may be difficult to solvate fully.

4 Selected Examples of Self-Assembling Systems

Many of the modules described above were designed to form closed or discrete aggregates. In some cases the modules simply dimerize or complex a complementary partner (see Figs. 1, 2, 5, and 8–14). The linking of modules to increase complex or dimer stability is a common theme in the area of self-assembly. For example, Hamilton has shown [26] that the K_{assoc} for $31 \cdot 34$, $4,500 \text{ M}^{-1}$, is about 10-fold higher than the analogous single base-pair (Fig. 20). A larger increase is predicted but strain in the macrocyclic complex was proposed to diminish the stability of $31 \cdot 34$. In a related study, Sessler reported that dimer $53 \cdot 53$ (Fig. 20) did not show appreciable dissociation upon dilution (^1H NMR monitoring) in chloroform-*d*, whereas complexes such as $40 \cdot 41$ do fully dissociate [25]. Most remarkable is dimer $77 \cdot 77$ (Fig. 12), which does not dissociate, even in the highly competitive solvent DMSO [52].

Whereas the hydrogen-bonding sites in the modules described above are roughly in parallel planes or slightly diverging, the ADA sites in ditopic receptor **35** converge and are designed to complex the intrinsic ditopic guest **38**. Thus, macrocyclic barbituric acid receptor **35** complexes diethyl barbituric acid (**38**) with a K_{assoc} of approximately 10^5 – 10^6 M^{-1} (Fig. 20) [77], a value that is 2–4 orders of magnitude higher than the DAD · ADA hydrogen-bonded complexes in Fig. 1. A 2 + 2 self-assembled structure, $(93)_2(94)_2$ (Fig. 21), reported by Lehn [78] takes advantage of the intrinsic ditopic DAD array in triaminotriazine module **93** and its complementarity to uracil.

Another common strategy for creating closed assemblies is the use of intrinsic ditopic modules in which two complementary hydrogen-bonding

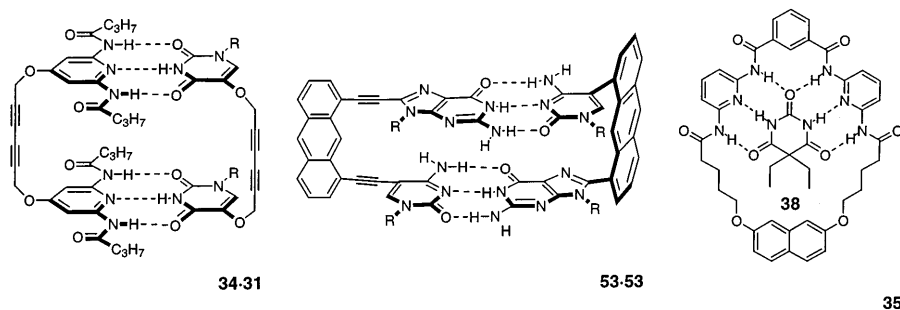


Fig. 20. Self-assembled structures involving dimeric modules

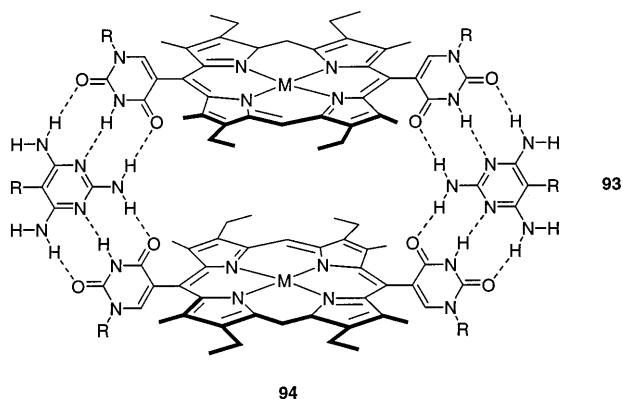


Fig. 21. A dimeric porphyrin via 2 + 2 self-assembly. R = decyl

sites are fixed at a 60° angle. As seen in Fig. 22, such modules are capable of forming cyclic hexamers. One of the most studied motifs uses analogs of two complementary, ditopic modules: cyanuric or barbituric acid and melamine. In this case, the hexamer $38 \cdot 95$ is formed from three molecules of **38** and three of **95**. This assembly (Ar = 4-*tert*-butylphenyl) is formed in the solid state, but the association in solution is weak [30, 79]. The structure is greatly stabilized, however, by covalently linking the melamine units, as in **36**, or the cyanuric acid modules, as in **39** (Fig. 4). An example of such an assembly is $(36)_3(38)_6$ (Fig. 22). This general strategy has been extensively investigated by both Whitesides [30, 79, 80] and Reinhoudt [29, 81].

The advantage of the melamine · cyanuric acid approach is the comparative ease with which these modules can be derivatized. Indeed, a multitude of different monomeric, dimeric, and trimeric modules have been synthesized and studied. The disadvantage of the approach is two-fold. The DAD · ADA motif is the weakest of the triply hydrogen-bonded arrays (*vide supra*), and its symmetry means that the components lack sufficient information to form the cyclic aggregate. Indeed, the same basic modules have been used to form polymeric aggregates (see below). To counter this problem, hexamers $(60)_6$ [42] and $(61)_6$ (R = C_7H_{15}) [43] were designed to form from a single, intrinsic ditopic module carrying both a DDA and AAD site. Module **60** forms a hexamer in solution, and **61** has been shown to form a hexamer both in solution [39, 42] and in the solid state [43] (Fig. 22). The advantage of using the AAD · DDA hydrogen-bonding motif is clear in that $(60)_6$ is sufficiently stable to form in competitive solvents such as aqueous tetrahydrofuran [42]. In a similar fashion, modules **57** and **62** were intended to pair with **59** and **63** (Fig. 7), respectively, to form 3 + 3 hexamers [40, 41]. In the latter case, however, the individual units are self-complementary and **62** crystallizes alone, forming a hydrogen-bonded ribbon in the solid state [40].

Larger superstructures can be formed by the subsequent aggregation of discrete assemblies or directly from self-assembling subunits that are designed to form polymeric aggregates. Along these lines, Kimizuka and Kunitake have

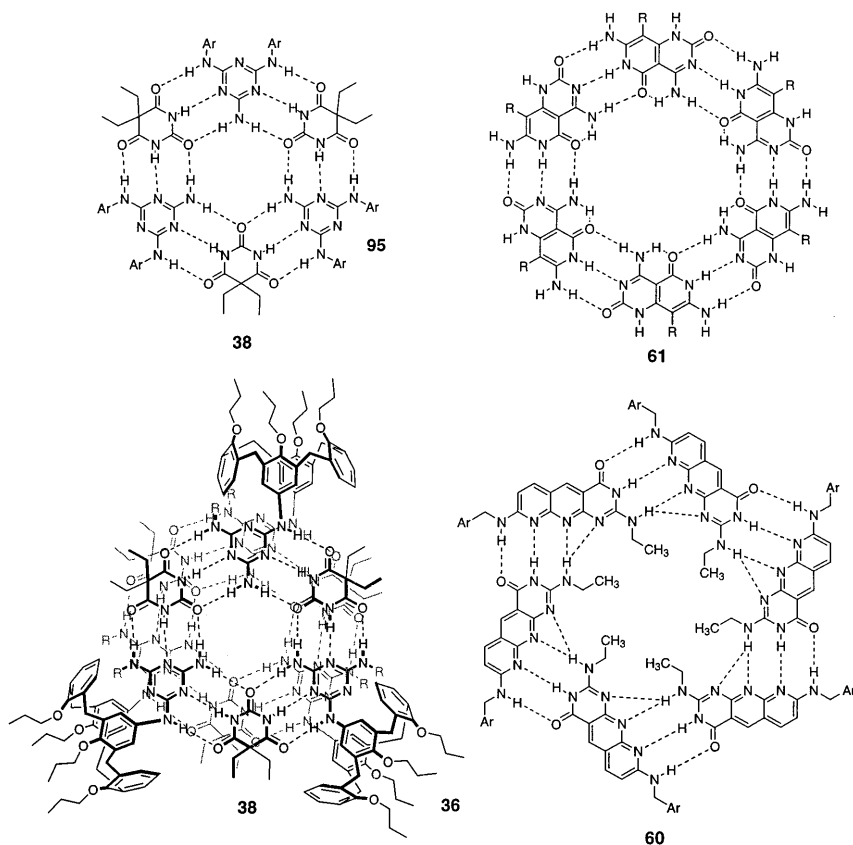


Fig. 22. Cyclic “hexameric” aggregates $(38)_3(95)_3$, $(60)_6$, $(61)_6$, and $(36)_3(38)_6$

studied [82] a 1:1 mixture of **37** and **96** ($R =$ long-chain dialkyl ether) and found evidence by IR and TEM for hydrogen-bonded assemblies appearing as strands with ca. 100 Å diameter (Fig. 23). Three models are consistent with such a structure: (1) tubes made by stacking $(37)_6(96)_6$, (2) helical tubes analogous to the schematic shown in Fig. 23, (3) zigzag ribbons of $37 \cdot 96$ packed face-to-face. Such zigzag ribbons have been a popular target for crystal engineers. In this regard, the DAD · ADA motif has been employed quite successfully. Lehn has investigated a barbituric acid · triaminotriazine system [83] and Whitesides has examined a cyanuric acid · melamine pair [79]. As noted above, cyclic structure $(38)_3(95)_3$ is favored by peripheral steric interactions with the *tert*-butylphenyl groups. When the *tert*-butyl groups are replaced with smaller groups, the 1:1 mixture crystallizes as a tape or “crinkled tape” [79]. Thus, as indicated above, the information content in the hydrogen-bonding system is insufficient to control which self-assembled structure is formed.

The DAD · ADA motif has also been used by Kunitake, Ringsdorf, and others for interfacial recognition and assembly [84, 85]. As seen in Fig. 24,

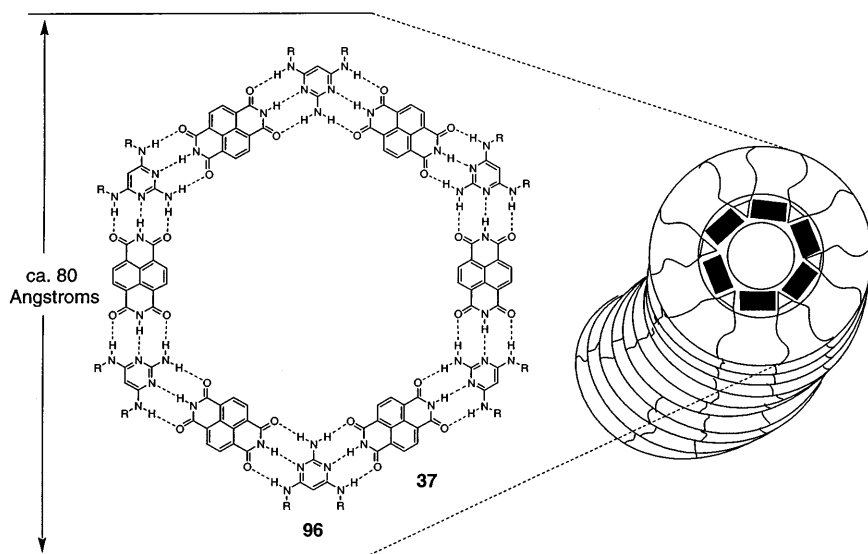


Fig. 23. One possible structure for the self-assembled material formed from a 1:1 mixture of 37 and 96. $R = C_{12}H_{25}OC_3H_6^-$ or $C_4H_9(C_2H_5)CH C_3H_6^-$

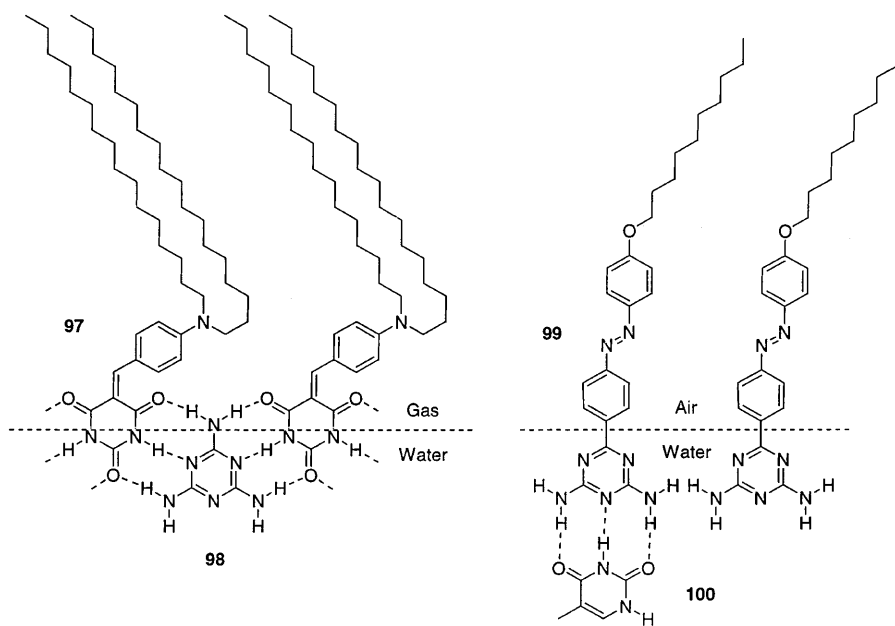


Fig. 24. Interfacial recognition and assembly at the gas-water interface using the DAD · ADA hydrogen-bonding motif

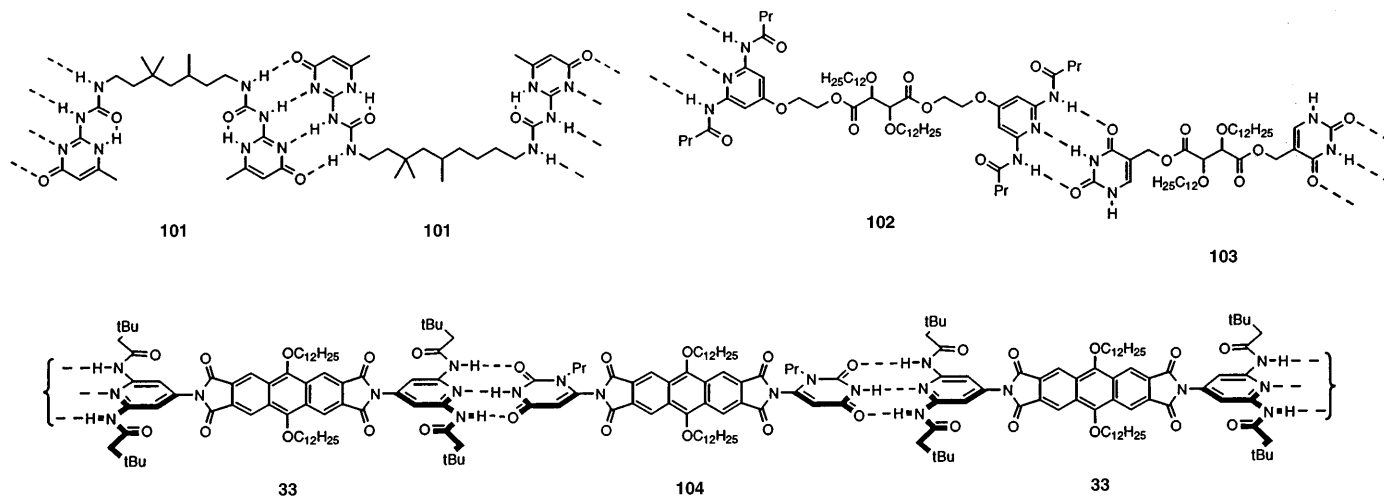


Fig. 25. Supramolecular polymers formed by hydrogen bonding in solution or mesophases

Ringsdorf reported [85] that melamine (**98**) organizes ditopic lipid **97** at the gas–water interface using the same polymeric hydrogen-bonding pattern found in the 1:1 crystals studied by Lehn and Whitesides (*vide supra*). Kunitake and co-workers studied the interfacial recognition of thymine **100** by a monolayer of amphiphilic diaminotriazine units at the air–water interface [85]. Normally, these base-pairing type interactions are entirely disrupted in aqueous media, but the microenvironment of organized assemblies allows this special recognition to occur [86].

Finally, there is considerable interest in polymeric assemblies both in solution and in liquid crystalline phases [87]. In a seminal report, Meijer and co-workers [49] have synthesized dimers of module **75** (e.g. **101**) and shown that its solutions have rheological properties similar to those shown by normal polymer solutions (Fig. 25). In this regard, the high dimerization constant of **75** allows a high degree of polymerization at accessible concentrations. Likewise, Lehn has shown that 1:1 mixtures of **102** · **103** and **33** · **104** form supramolecular, polymeric, liquid crystalline phases (Fig. 25). The structure of **102** · **103** is believed to contain a triple helical superstructure [88], whereas rigid assembly **33** · **104** forms a lyotropic mesophase [89].

5

Conclusions

A broad set of heteroaromatic modules is now available for constructing self-assembling systems. Some of the modules are trivial to prepare while others require rather lengthy syntheses. There is also considerable variability in the strength with which these modules self-associate or bind their complement, which can be useful for fine-tuning the stability of self-assembled structures. Many factors contribute to this variability, making the a priori prediction of complex stability a daunting task. Nonetheless, when one considers the number of primary and secondary hydrogen-bonding interactions in the complexes, as well as the degree of preorganization of complementary modules, most differences in complex strengths can be explained.

The use of these modules in self-assembling systems has been extremely successful, as evidenced by the few examples presented in Sect. 4. Nature has chosen base-pairing as a means of storing genetic information because of the reversibility of hydrogen bonding. Similarly, applications of the modules described herein must benefit from the reversible nature of the self-assembly process. Thus, as the complexity of the self-assembly processes increases, one obvious advantage of the hydrogen-bonding approach is the ability of such systems to automatically edit out kinetically accessible assembly errors.

References

1. (a) Breslow R (1998) *Chem Biol* 5: R27; (b) Breslow R (1995) *Acc Chem Res* 28: 146
2. Eigen M, De Maeyer L (1966) *Naturwissenschaften* 53: 50
3. Lehn JM (1988) *Angew Chem* 27: 89

4. Zimmerman SC, Murray TJ (1994) In: Wipff G (ed) Computational approaches in supramolecular chemistry. NATO ASI Series, vol 426. Kluwer, Amsterdam, p 109
5. Kyogoku Y, Lord RC, Rich A (1967) Proc Natl Acad Sci USA 57: 250
6. Bell TW, Hou Z, Zimmerman SC, Thiessen PA (1995) Angew Chem Int Ed Engl 34: 2163 and references cited therein
7. Hamilton AD, Pant N, Muehldorf A (1988) Pure Appl Chem 60: 533
8. Rebek J Jr (1990) Angew Chem Int Ed Engl 29: 245
9. Zimmerman SC, Wu W (1989) J Am Chem Soc 111: 8054
10. Kelly TR, Maguire MP (1987) J Am Chem Soc 109: 6549
11. Kirnos MD, Khudyakov IY, Alexandrushikina NI, Vanyushin BF (1997) Nature 270: 369
12. Gryaznov S, Schultz RG (1994) Tetrahedron Lett 35: 2489 and references cited therein
13. Haaima G, Hansen HF, Christensen L, Dahl O, Nielsen PE (1997) Nucleic Acids Res 25: 4639
14. Hamilton AD, Van Engen D (1987) J Am Chem Soc 109: 5035
15. Park TK, Schroeder J, Rebek J Jr (1991) J Am Chem Soc 113: 5125
16. Schneider HJ, Juneva RK, Simova S (1989) Chem Ber 122: 1211
17. Kelly TR, Bridger GJ, Zhao C (1990) J Am Chem Soc 112: 8024
18. Murray TJ, Zimmerman SC (1992) J Am Chem Soc 114: 4010
19. Beijer FH, Sijbesma RP, Vekemans JAJM, Meijer EW, Kooijman H, Spek AL (1996) J Org Chem 61: 6371
20. Adrian JC Jr, Wilcox CS (1991) J Am Chem Soc 113: 678
21. Murray TJ, Zimmerman SC, Kolotuchin SV (1994) Tetrahedron 51: 635 and references cited therein
22. Feibush B, Figueroa A, Rosita C, Onan KD, Feibush P, Karger BL (1986) J Am Chem Soc 108: 3310
23. Hamilton AD, Van Engen D (1987) J Am Chem Soc 109: 5035
24. Leonard NJ, McCredie RS, Logue MW, Cunddall RL (1973) J Am Chem Soc 95: 2320
25. Sessler JL, Wang R (1998) J Org Chem 63: 4079
26. Hamilton AD, Little D (1990) J Chem Soc Chem Commun 297
27. Schall OF, Gokel GW (1994) J Am Chem Soc 116: 6089
28. Kotera M, Lehn J-M, Vigneron JP (1995) Tetrahedron 51: 1953
29. Prins LJ, Huskens J, de Jong F, Timmerman P, Reinhoudt DN (1999) Nature 398: 498
30. Mammen M, Simanek EE, Whitesides GM (1996) J Am Chem Soc 118: 12614
31. Berstein J, Sterns B, Shaw E, Lott WA (1947) J Am Chem Soc 69: 1151
32. (a) Berman A, Izraeli ES, Levanon H, Wang B, Sessler JL (1995) J Am Chem Soc 117: 8252; (b) Sessler JL, Wang B, Harriman A (1993) J Am Chem Soc 115: 10418
33. Roberts C, Bandaru R, Switzer C (1997) J Am Chem Soc 119: 4640
34. Kral V, Sessler JL (1995) Tetrahedron 51: 539
35. Hisatome M, Ikeda K, Kishibata S, Yamakawa K (1993) Chem Lett 1357
36. Chen L, Sakai N, Moshiri ST, Matile S (1998) Tetrahedron Lett 39: 3627
37. Fenlon EE, Murray TJ, Baloga MH, Zimmerman SC (1993) J Org Chem 58: 6625
38. Marsh A, Nolen EG, Gardinier KM, Lehn J-M (1994) Tetrahedron Lett 35: 397
39. Marsh A, Silvestri M, Lehn J-M (1996) Chem Commun 1527
40. Lehn J-M, Mascal M, DeCian A, Fischer J (1992) J Chem Soc Perkin Trans 2: 461
41. Petersen P, Wu W, Fenlon EE, Kim S, Zimmerman SC (1996) Bioorg Med Chem 4: 1107
42. Kolotuchin SV, Zimmerman SC (1998) J Am Chem Soc 120: 9092
43. Mascal M, Hext NM, Warmuth R, Moore MH, Turkenburg JP (1996) Angew Chem Int Ed Engl 35: 2204
44. Meyer H, Bossert F, Horstmann H (1978) Liebig's Ann Chem 1476
45. Bell DA, Anslyn EV (1995) Tetrahedron 51: 7161
46. Beijer FH, Kooijman H, Spek AL, Sijbesma RP, Meijer EW (1998) Angew Chem Int Ed Engl 37: 75
47. Corbin PS, Zimmerman SC (1998) J Am Chem Soc 120: 9710
48. Beijer FH, Sijbesma RP, Kooijman H, Spek AL, Meijer EW (1998) J Am Chem Soc 120: 6761

49. Sijbesma RP, Beijer FH, Brunsveld L, Folmer BJB, Hirschberg JHK, Lange RFM, Lowe JKL, Meijer EW (1997) *Science* 278: 1601
50. Pfeleiderer M, Pfeleiderer W (1992) *Heterocycles* 33: 905
51. Corbin PS, unpublished results
52. Sessler JL, Wang R (1998) *Angew Chem Int Ed Engl* 37: 1726
53. Lüning U, Köhl C (1998) *Tetrahedron Lett* 39: 5735
54. Corbin P, Zimmerman SC, submitted for publication
55. Jorgensen WL (1998) *J Phys Chem* 102: 3782
56. Hopkins HP, Alexander CJ, Ali SZ (1978) *J Phys Chem* 82: 1268
57. Schneider HJ, Juneva RK, Simova S (1989) *Chem Ber* 122: 1211. This analysis has been updated to include secondary hydrogen bonding (see [61])
58. Jorgensen WL, Pranata J (1990) *J Am Chem Soc* 112: 2008; Pranata J, Wierschke SG, Jorgensen WL (1991) *J Am Chem Soc* 113: 2810
59. Zimmerman SC, Murray TJ (1994) *Tetrahedron Lett* 35: 4077
60. Gardner RR, Gellman SH (1995) *J Am Chem Soc* 117: 10411
61. Sartorius J, Schneider HJ (1996) *Eur J Chem* 2: 1446
62. Cram DJ (1986) *Angew Chem Int Ed Engl* 25: 1039
63. Cram DJ, Trueblood KN (1985) in Vögtle F, Weber E (eds) *Host guest complex chemistry, macrocycles*. Springer, Berlin Heidelberg New York, p 125
64. Zimmerman SC, Mrksich M, Baloga M (1989) *J Am Chem Soc* 111: 8528
65. Page MI, Jencks WP (1971) *Proc Natl Acad Sci USA* 68: 1678
66. Searle MS, Williams DH (1992) *J Am Chem Soc* 114: 10690
67. Eblinger F, Schneider HJ (1998) *Angew Chem Int Ed Engl* 37: 826. This paper, which claims to report the first study wherein the number of free rotations within supramolecular complexes are systematically varied (see however [64]), studies the interaction between dicarboxylates and diamide. It is complicated by the fact that rigid and flexible dicarboxylates may bind the diamides in different orientations
68. Hamilton AD, Pant N (1998) *J Chem Soc Chem Commun* 765
69. Murray TJ, Zimmerman SC (1995) *Tetrahedron Lett* 36: 7627
70. Watson JD, Crick FHC (1953) *Nature* 171: 964
71. Douhal A, Kim SK, Zewail AH (1995) *Nature* 378: 260
72. Robinson H, Gao YG, Bauer C, Roberts C, Switzer C, Wang AHJ (1998) *Biochemistry* 37: 10897
73. Goodman MF (1995) *Nature* 378: 237
74. Beak P (1977) *Acc Chem Res* 10: 186
75. Fersht AR (1987) *Trends Biochem Sci* 12: 301
76. Fan E, van Araman S, Kincaid S, Hamilton AD (1993) *J Am Chem Soc* 115: 369; Wilcox CS, Kim EI, Romano D, Kuo LH, Burt AR, Curran DP (1995) *Tetrahedron* 55: 621
77. Hamilton AD, Fan E, Van Arman S, Vicent C, Tellado FG, Geib SJ (1993) *Supramol Chem* 1: 247
78. Drain CM, Fischer R, Nolen EG, Lehn J-M (1993) *J Chem Soc Chem Commun* 243
79. Zerkowski JA, Seto CT, Whitesides GM (1992) *J Am Chem Soc* 114: 5473
80. Seto CT, Whitesides GM (1993) *J Am Chem Soc* 115: 905
81. Crego Calama M, Fokken R, Nibbering NMM, Timmerman P, Reinhoudt DN (1998) *J Chem Soc Chem Commun* 1021
82. Kimizuka N, Kawasaki T, Hirata K, Kunitake T (1995) *J Am Chem Soc* 117: 6360
83. Lehn J-M, Mascal M, DeCian A, Fischer J (1990) *J Chem Soc Chem Commun* 479
84. Ariga K, Kunitake T (1998) *Acc Chem Res* 31: 371 and references cited therein
85. Bohanon TM, Denzinger S, Fink R, Paulus W, Ringsdorf H, Weck M (1995) *Angew Chem* 34: 58
86. Nowick JS, Cao T, Noronha G (1994) *J Am Chem Soc* 116: 3285
87. Zimmerman N, Moore JS, Zimmerman SC (1998) *Chem Ind* 604
88. Gulick-Krymicki T, Fouquey AM, Lehn J-M (1993) *Proc Natl Acad Sci* 90: 163
89. Kotera M, Lehn J-M, Vigneron JP (1994) *J Chem Soc Chem Commun* 197