Reprinted from ACCOUNTS OF CHEMICAL RESEARCH, 1995, 28.

Noncovalent Synthesis: Using Physical–Organic Chemistry To Make Aggregates

GEORGE M. WHITESIDES,* ERIC E. SIMANEK, JOHN P. MATHIAS, CHRISTOPHER T. SETO, DONOVAN N. CHIN, MATHAI MAMMEN, AND DANA M. GORDON

Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received April 18, 1994

What Kinds of Structures Should Organic Chemists Be Learning To Make?

A *molecule* is usually understood to be a stable collection of atoms connected by a continuous network of covalent bonds. The development of methods for constructing these networks has been a central occupation of organic chemistry, and the success of these methods has made possible the power, elegance, and utility of modern organic synthesis. The preparations of vitamin B_{12},[1] palytoxin,[2] calicheamicin,[3] and other complex secondary metabolites illustrate the extraordinary sophistication of this field. This type of synthesis—which we refer to as *covalent* synthesis, in the absence of a better term— continues to expand its capabilities, but it may be understandably difficult to provide very large and structurally complex molecules quickly and economically by using it.[4]

Organic chemistry has always taken much of its inspiration and motivation from Nature. As biological molecules—especially large molecules having complex tertiary structures such as proteins, DNA, and RNA—have become central concerns of organic chemistry, *noncovalent* interactions have moved toward the center of attention. Although biological macromolecules are largely composed of covalent bonds, the networks of these bonds are *not* always continuous, and many important structures—including multimeric proteins and DNA itself—are "aggregates" and not simply "molecules". Many biological molecules and aggregates derive much of their unique structure and function from noncovalent interactions: that is, from hydrophobic and ionic interactions and hydrogen bonds. The formation of a phospholipid bilayer, noncovalent association of a protein and ligand, interaction of a transcription factor with DNA, and folding of a tRNA into its three-dimensional conformation are all examples of processes that depend on noncovalent interactions.

Nonbiological and biological organic chemistries place different importance on noncovalent interactions: for biological chemistry, they are of primary importance; for nonbiological chemistry, they are secondary (Table 1). Many of the most important issues in biochemistry hinge on understanding and controlling noncovalent interactions. We, and others, believe that the universal importance of noncovalent interactions in the assembly and stabilization of biological systems provides a compelling motivation to consider a new type of organic synthesis focused on these interactions. They are also critically important in materials science, although their applications in this area involves a different set of issues than those for soluble aggregates.

Understanding noncovalent bonding sufficiently to use it in the design and preparation of new chemical entities having the structural and functional complexity of biological macromolecules requires a fundamentally different intellectual and technical strategy than that used in covalent, nonbiological synthesis. In covalent synthesis, bonds are characterized primarily by their enthalpy: in general, bonds and products, once formed, are kinetically stable (although they may be formed by processes that are reversible). Concepts such as bond energy, strain, and stereoelectronic interaction provide the vocabulary for discussions. Yields are often substantially less than quantitative—a characteristic that becomes a concern as the number of steps in a synthetic pathway increases. In noncovalent synthesis, the products are equilibrating structures. They reflect a *balance* between enthalpy and entropy: potential products of these syntheses must be evaluated in terms of thermodynamic minima in equilibrating mixtures. The major challenge is predicting and then influencing the position of these equilibria to direct association toward a specific target/aggregate.

Are there alternatives to covalent synthesis for the preparation of large, complex, but well-defined struc-

George M. Whitesides was born in Louisville, KY, in 1939. He received his B.A. from Harvard in 1960 and his Ph.D. with John D. Roberts from the California Institute of Technology in 1964. After spending almost 20 years at MIT, he joined the Department of Chemistry at Harvard University in 1982. His research interests include materials science, surface chemistry, rational drug design, molecular virology, and molecular recognition.

Eric E. Simanek was born in Tuscola, IL, in 1969. He obtained his B.S. at the University of Illinois at Urbana–Champaign in 1991. He is currently a doctoral candidate with G.M.W.

John P. Mathias was born in Manchester, England, in 1966. After obtaining his B.Sc. at the University of Sheffield and staying on to complete his Ph.D. with J. F. Stoddart, he completed a NATO-SERC post-doctoral fellowship with G.M.W. He is currently with Pfizer Central Research, England.

Christopher T. Seto was born in Seattle, WA, in 1964 and obtained his Ph.D. with G.M.W. in 1992 as an Eli Lilly Predoctoral Fellow. He was a post-doctoral fellow with P. Bartlett at the University of California at Berkeley and is currently an assistant professor at Brown University.

Donovan N. Chin was born in Croydon, England, in 1964. He received his B.Sc. at the University of Florida in 1987 and obtained his Ph.D. at the University of Massachusettes at Lowell with A. C. Watterson. He is currently a post-doctoral fellow with G.M.W.

Mathai Mammen was born in Vellore, India, in 1967. He obtained his B.Sc. at Dalhousie University in 1989. He is currently an Eli Lilly predoctoral fellow with G.M.W.

Dana M. Gordon was born in 1962 in Santa Monica, CA. After completing his B.S. at the University of California at Los Angeles and Ph.D. at Yale University with S. Danishefsky, he was an NIH post-doctoral fellow with G.M.W. He is currently an assistant professor at Brandeis University.

(1) (a) Woodward, R. B. *Pure Appl. Chem.* **1973**, *33*, 145. (b) Eschenmoser, A. E.; Wintner, C. *Science* **1977**, *196*, 1410.
(2) Kishi, Y. *Pure Appl. Chem.* **1989**, *61*, 313.
(3) Groneberg, R. D.; Miyazaki, T.; Stylianides, N. A.; Schulze, T. J.; Stahl, W.; Schreiner, E. P.; Suzuki, T.; Iwabachi, Y.; Smith, A. L.; Nicolaou, K. C. *J. Am. Chem. Soc.* **1993**, *115*, 7593.
(4) The synthesis of polynucleotides, polypeptides, and polysaccharides produces large molecules.

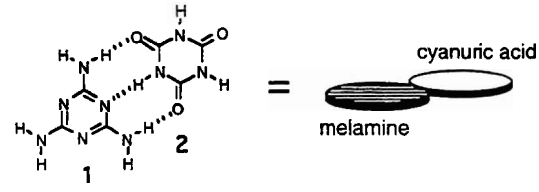
0001-4842/95/0128-0037$09.00/0

Table 1. Comparison of Covalent and Noncovalent Synthesis

	covalent	noncovalent
constituent bond types	covalent	ionic, hydrophobic, hydrogen bonds
bond strengths (kcal/mol)	25–200	0.1–5
stability of bonds in the product	kinetically stable	kinetically reversible
contributions to ΔG	usually dominated by ΔH	ΔH and $T\Delta S$ are often comparable
strategy of design	selective *reaction*	directed *equilibrium*
importance of solvent effects	secondary	primary
other characteristics		cooperative behavior important
stimuli	secondary metabolites	proteins, nucleic acids

tures that combine flexibility with practicality? There are many routes to macromolecules, but each has limitations. Covalent polymerization can generate very high molecular weights (for example, polyolefins), but does not offer a high degree of flexibility in the control of molecular structure. The specialized, repetitive strategies used in chemical synthesis of polypeptides, oligonucleotides, and oligosaccharides are also restricted in the types of structures they can generate. The routes used to synthesize dendrimers offer a compromise between covalent synthesis and covalent polymerization, but again with limitations in the extent of control over the product structures.[5] Biotechnology and enzymology offer excellent routes, but are most useful for proteins, nucleic acids, oligosaccharides, and other water-soluble structures. We are interested in the idea that noncovalent assembly of small molecules may enable synthetic chemists to make aggregates of the same degree of structural complexity and, perhaps ultimately, functionality as highly structure biomolecules, but with fewer (or at least different) restrictions on the types of structures that can be generated. This approach to synthesis will require a higher level of understanding and control of the thermodynamics of aggregation phenomena—and especially of the entropic features of noncovalent association—than is currently practiced in synthetic organic chemistry.

To demonstrate concepts in noncovalent synthesis, we have focused on a model system—aggregates derived from the cyanuric acid (**1**)·melamine (**2**) lattice (CA·M)[6] (Figure 1)—in organic solution. The CA·M lattice is remarkably stable: it can be heated at 350 °C without decomposition.[7] This stability reflects, we presume, the fact that removing a molecule of CA or M from the lattice requires breaking nine hydrogen bonds, each with energies close to 7 kcal/mol.[8] The CA·M lattice provides the conceptually simplest model system that we can envision for studying *non*covalent synthesis. Of course, studies centered on it benefit enormously from the wealth of superb prior work on aggregates held together by hydrogen bonds from the laboratories of Rebek,[9] Hamilton,[10] Wilcox,[11] Zimmerman,[12] and Wuest[13] and from the work of others—Lehn,[14] Pedersen,[15] Cram,[16] Stoddart,[17] Diederich,[18] Sauvage,[19] Breslow,[20] Still,[21] and Kunitake[22]—who have delineated the fundamental features of molecular recognition. Excellent reviews[23] describe the potential of biological paradigms based on molecular self-



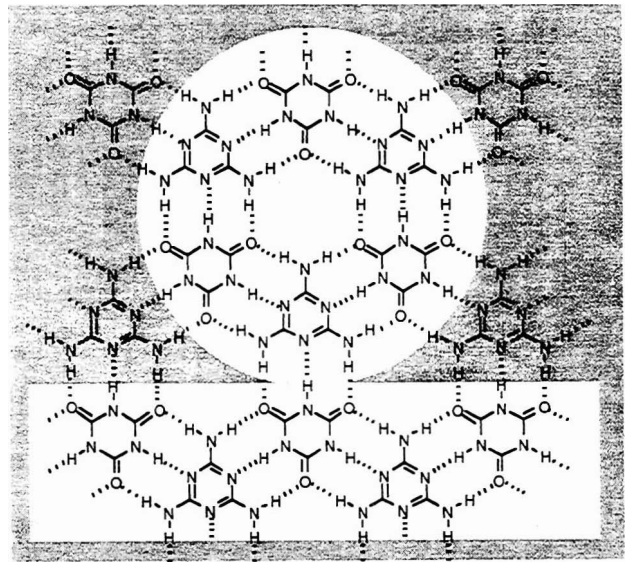


Figure 1. Melamine (**1**, dark disk) and cyanuric acid (**2**, white disk) lattice (CA·M). Two of the possible structural motifs that can be derived from the lattice—a cyclic rosette and a linear tape—are indicated. A third motif—a crinkled tape—(not shown) is also observed in the solid state. Other patterns are too congested sterically to exist under most circumstances.

(5) See, for example: Xu, Z.; Moore, J. S. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1354.
(6) X-ray structure of CA·M·3HCl: Wang, Y.; Wei, B.; Wang, Q. J. *Crystallogr. Spectrosc. Res.* **1990**, *20*, 79.
(7) Tsuya, Y.; Watanabe, M.; Hirae, T.; Sato, M.; Kawakita, A. *ASLE Trans.* **1981**, *24*, 49.
(8) A range of values for the enthalpy of a hydrogen bond have been reported. For CA·M·HCl: Finkel'shtein, A. I.; Rukevich, O. S. *Zh. Prikl. Spectrosck.* (*Russ.*) **1983**, *38*, 327. For a recent review on hydrogen bonding in the solid state and references for hydrogen bond enthalpies in these systems see Aakeroy, C. B.; Seddon, K. R. *Chem. Soc. Rev.* **1993**, 397.
(9) Wyler, R.; de Mendoza, J.; Rebek, J. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1699.
(10) Geib, S. J.; Vicent, C.; Fan, E.; Hamilton, A. D. *Angew. Chem., Int. Ed. Engl.* **1993**, *105*, 80.
(11) Webb, T. H.; Suh, H.; Wilcox, C. S. *J. Am. Chem. Soc.* **1991**, *113*, 8554.
(12) Zimmerman, S. C.; Saionz, K. W.; Zeng, Z. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1190.
(13) Simard, M.; Su, D.; Wuest, J. D. *J. Am. Chem. Soc.* **1991**, *113*, 4696.
(14) (a) Baxter, P.; Lehn, J. M.; DeCian, A.; Fisher, J. *Angew. Chem., Int. Ed. Engl.* **1993**, *28*, 89. (b) Lehn, J. M. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 89.
(15) Pedersen, C. J. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 1021, and references cited therein.
(16) (a) Helgeson, R. C.; Selle, B. J.; Goldberg, I.; Knobler, C. B.; Cram, D. J. *J. Am. Chem. Soc.* **1993**, *115*, 11506. (b) Cram, D. J. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 1009.
(17) Ballardini, R.; Balzani, V.; Gandolfi, M. T.; Prodi, L.; Venturi, M.; Philp, D.; Ricketts, H. G.; Stoddart, J. F. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1301.
(18) Jorgenson, W. L.; Nguyen, T. B.; Sanford, E. M.; Chao, I.; Houk, K. N.; Diederich, F. *J. Am. Chem. Soc.* **1992**, *114*, 4003.
(19) Nierengarten, J. F.; Dietrich-Buchecker, C. O.; Sauvage, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 375.
(20) Breslow, R.; Graft, A. *J. Am. Chem. Soc.* **1993**, *115*, 10988.
(21) Borchardt, A.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 373.
(22) Kunitake, T. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 709.

assembly as a strategy for synthesis and the use of hydrogen bonds in such structures.

In brief, we wish to (i) define strategies for the assembly of supramolecular aggregates based on the CA·M lattice, (ii) develop techniques for the complete characterization and study of these entities, and (iii) design models to predict and assess the relative stabilities of these noncovalently-bonded structures. Understanding these three issues will help in using molecular self-assembly as a strategy in chemical synthesis, in materials science, and in molecular biology. We emphasize that the focus of this work is on *synthesis* and in understanding how to incorporate concepts and relatively simple structures from molecular recognition into a set of procedures for making structurally complex aggregates.

Nomenclature. The aggregates that are the products of these syntheses are structurally too complex to be represented conveniently in full atomic detail. The system we use to abbreviate their structures is summarized in Figure 1. Melamines are represented by dark disks. Derivatives of cyanuric acid (formally described as isocyanuric acids and barbituric acids) are indicated by white disks. We refer to the cyclic aggregate of three melamine and three isocyanuric acids as a "rosette". Other details of molecular structure are provided in whatever detail is required to make the points being discussed.

Design. Aggregates based on the CA·M rosette are held together by 18 hydrogen bonds. In chloroform solution, these bonds have enthalpies of 1–3 kcal/mol.[24] The total enthalpy of these bonds is, thus, substantial but is counterbalanced by a number of unfavorable entropic terms due to losses in freedom in translation and rotation of the components on formation of the aggregate. The maximum enthalpy of formation of the network of hydrogen bonds is fixed for a single CA·M rosette. In chloroform solution, we estimate the enthalpy of the network to be approximately 24 kcal/mol.[25] A single rosette is an aggregate of six molecules and is strongly disfavored entropically. To favor the formation of a rosette, we must minimize the unfavorable contributions of translational and rotational entropy. We have used two strategies for designing aggregates of the CA·M lattice: *preorganization*[16,26] (a qualitative concept that has become a foundation of molecular recognition) and *peripheral crowding*[27] (Figure 2). Both strategies are successful in forming stable aggregates; those based on preorganization have so far produced the more stable aggregates.

Preorganization. Linking three components of a rosette (usually the melamines since they are easier to synthesize than the isocyanurates) to a common hub by appropriate connecting groups "preorganizes" them for aggregation.[16,26] This strategy reduces the unfavorable changes in *translational* and *rotational* entropies that accompany the formation of the aggregate.

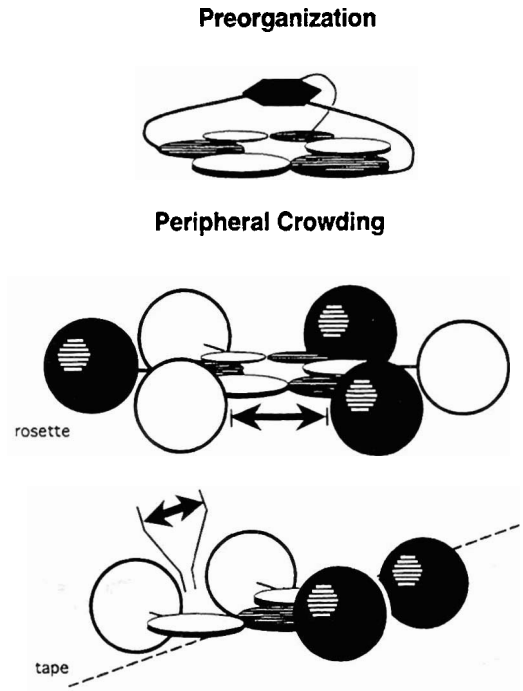


Figure 2. Preorganization describes the connection of melamines (or isocyanurates) using covalent bonds in a way that minimizes the unfavorable entropy of forming an aggregate. Peripheral crowding describes the use of bulky groups that favor rosette motifs over tape motifs. The arrows indicate regions of potential steric overlap.

Peripheral Crowding. Choosing the substituents on the melamine and isocyanuric acid groups in a way that makes noncyclic tapes (Figure 2) energetically unfavorable promotes the formation of cyclic rosettes.[27] This manipulation of conformational enthalpy rests on the assumption that the melamine and isocyanuric acid derivatives interact sufficiently strongly that they will normally form 1:1 aggregates of some structure: the problem is to favor rosettes over tapes.

Synthesis. Synthesis of the aggregates often involves no more than mixing the components in chloroform. In instances in which a component is insoluble, it can be useful to dissolve both components in methanol or some more polar solvent, remove this solvent, and redissolve the residue in chloroform: this procedure often overcomes kinetic limitations due to solubilities.

Figure 3 summarizes schematically a number of the aggregates that we have made.[28] These structures are arranged qualitatively in order of stability—from lowest (top) to highest (bottom)—together with a parameter HB/(N − 1) that estimates stability based on techniques described in the next section.

(23) (a) Lindsey, J. S. *New J. Chem.* **1991**, *15*, 153. (b) Etter, M. C. *J. Phys. Chem.* **1991**, *95*, 4601. (c) Schneider, H. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 1417.

(24) (a) Gelman, S. H.; Dado, G. P.; Liang, G. B.; Adams, B. R. *J. Am. Chem. Soc.* **1991**, *113*, 1164. (b) Nikolic, A. D.; Tarkani-Rozsa, M.; Perisic-Janjic, N. U.; Ptrik, A.; Antonovic, D. G. *J. Mol. Struct.* **1990**, *219*, 245. (c) Williams, L. D.; Chawla, B.; Shaw, B. R. *Biopolymers* **1987**, *26*, 591.

(25) This value comes from exchange experiments detailed in Seto, C. T.; Whitesides, G. M. *J. Am. Chem. Soc.* **1993**, *115*, 1330.

(26) For a more thorough discussion of preorganization in our system, see Seto, C. T.; Whitesides, G. M. *J. Am. Chem. Soc.* **1993**, *115*, 905.

(27) For a more thorough discussion of peripheral crowding in the solution state, see Mathias, J. P.; Simanek, E. E.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 4326. In the solid state, see Zerkowski, J. A.; Mathias, J. P.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 4305, and references cited therein.

(28) Aggregate **3**: Mathias, J. P.; Simanek, E. E.; Zerkowski, J. A.; Seto, C. T.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 4316. Aggregate **4**: See ref 26. Aggregate **5**: See ref 27. Aggregate **6**: Seto, C. T.; Whitesides, G. M. *J. Am. Chem. Soc.* **1991**, *112*, 6409. Aggregate **7**: Mathias, J. P.; Seto, C. T.; Simanek, E. E.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 1725. Aggregate **8**: Mathias, J. P.; Simanek, E. E.; Seto, C. T.; Whitesides, G. M. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1766. Aggregate **9**: Seto, C. T.; Whitesides, G. M. *J. Am. Chem. Soc.* **1991**, *112*, 712. Seto, C. T.; Mathias, J. P.; Whitesides, G. M. *J. Am. Chem. Soc.* **1993**, *115*, 1321. Aggregate **10**: Reported with **7**. Aggregate **11**: Reported with **7**. Aggregate **12**: See ref 25.

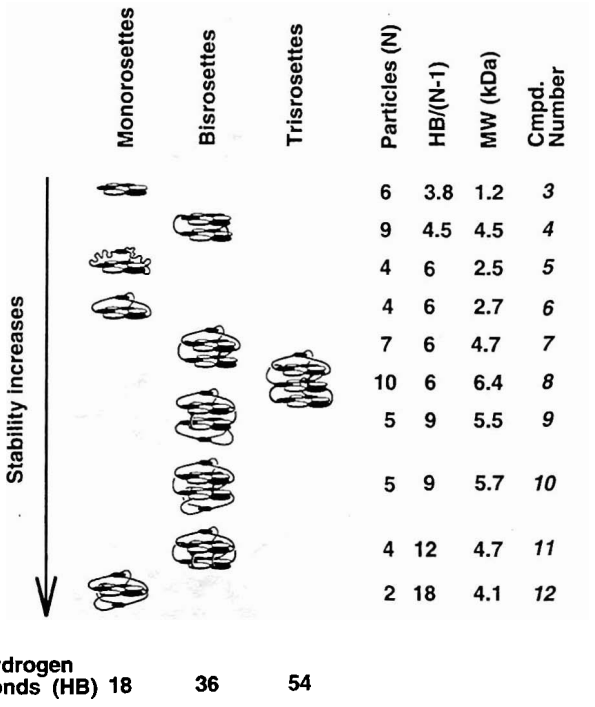


Figure 3. Structures based on the CA·M lattice, arranged in approximate order of stability. These structures are also organized by the number of rosettes (designated mono-, bis-, and tris-) they incorporate. Only one conformational isomer of each aggregate is shown; the stable conformer is, in general, not known. "Particles" is the number (N) of separate molecules comprising the aggregates—the larger the value, the greater the unfavorable entropic cost of aggregation. HB is the number of hydrogen bonds in the aggregates—the larger the value, the greater the favorable enthalpic gain of aggregation. HB/(N − 1) is an empirical parameter incorporating these trends; it is discussed in more detail in the text. MW is molecular weight of the aggregate. Aggregates discussed in the text are identified by compared number, (Cmpd. Number).

Characterization. The characterization of noncovalent aggregates has been a challenge that is still only partially met. A number of the techniques most useful in characterizing covalent structures—especially mass spectroscopy and X-ray crystallography—have not been useful with these aggregates.[29] Instead, characterization is based on combining inferences from a set of four methods. Individually, each method gives incomplete information; collectively, they can make a strong case.

Titration: Solubilization of Insoluble Components and Resolution of the NMR Spectrum. The aggregates are often much more soluble than one or both of the starting materials. Simple alkyl isocyanuric acids, for example, are almost insoluble in chloroform. Mixing 1 equiv of the trismelamine **13** with 3 equiv of the insoluble isocyanurate **14** results in complete dissolution of **14** (Figure 4). If additional **14** is added, it does not dissolve. This qualitative observation provides good evidence that a complex of **13**:**14** having the stoichiometry 1:3 is formed, but provides no further evidence concerning its structure.

Titration experiments monitored by NMR (^{1}H or ^{13}C) corroborate a 1:3 stoichiometry of **13**:**14** (Figure 5). The broadened, nondescript spectrum of **13** in solution is due to conformational isomers and intermolecular aggregation. It is transformed to a well-defined spectrum (superimposed on the original broadened spectrum) on addition of 1 equiv of **14**. Additional **14** leads to increasing intensity of the defined spectrum until the 1:3 stoichiometry is reached. Further addition of **14** does not effect the intensity or sharpness of the spectrum (Figure 5); the added **14** does not go into solution.

Vapor Pressure Osmometry. Classical colligative techniques for the determination of molecular weight give a semiquantitative, average measure of this property. We have used vapor pressure osmometry (VPO) to estimate their molecular weight (MW), because the aggregates dissociate during attempted mass spectrometry. Interpretation of the data from VPO requires calibration of the experiment using standards and interpretations of nonidealities. Using the standards we have selected, the values of MW inferred experimentally for the aggregates are typically within 20% of the theoretical values.

Gel Permeation Chromatography. GPC can establish that the aggregate exists as a single species and can give a qualitative estimate of the stability of this species. This technique separates species by size and would be expected to show separate peaks for **13**, **14**, **6**, and higher aggregates. In fact, using chloroform as solvent, neither **13** nor **14** emerge from a polystyrene GPC column at all, while the aggregate **6** emerges as a single peak with the retention time expected for an aggregate with a stoichiometry of **13**·**14**$_3$ (Figure 6). The *shape* of the GPC peak also provides information. If the complex dissociates on the column into components that migrate with different rates, then the peak tails. That for **3** does tail, but more stable complexes such as **6** tail much less. The order of stability inferred from GPC agree qualitatively with that estimated from HB/(N − 1).[30]

NMR Spectroscopy. For the simpler aggregates (e.g., **6**) it is possible to assign all the peaks in the ^{1}H NMR spectrum and to show that the details of the spectrum are those expected. Nuclear Overhauser experiments are especially useful for establishing the connectivity between molecules of CA and M within the rosette. Methylene groups show the diastereotopic protons expected for a chiral environment in the aggregate. For the larger complexes, most of the proton spectrum is uninterpretable.

Fortunately, the region between 16 and 13 ppm of the ^{1}H spectra of these aggregates contains only peaks due to the N−H protons of the isocyanuric acid or barbituric acid molecules. Figure 7 shows representative spectra for **6** (when X = CEt_2; Figure 4). At room temperature, this part of the spectrum shows the time-averaged symmetry of the aggregate; as the temperature is lowered, new peaks appear. These peaks can readily be rationalized on the basis of an equilibrium between one isomer with C_3 symmetry and a second isomer with C_1 symmetry. In the former species, the N−H protons occur in only two different environments; in the latter, all six protons are, in principle and often in practice, distinguishable.[31]

(29) For crystal structures of tapes and of a discrete rosette, see Zerkowski, J. A.; MacDonald, J. C.; Seto, C. T.; Wierda, D. A.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 2382, and references cited therein.

(30) The retention time of an aggregate by GPC correlates with ln(molecular weight) for a variety of aggregates based on the CA·M lattice. For a more thorough discussion of this observation, see ref 28, aggregate **7**.

(31) Simanek, E. E.; Wazeer, M. I. M.; Mathias, J. P.; Whitesides, G. M. *J. Org. Chem.* **1994**, *59*, 4904.

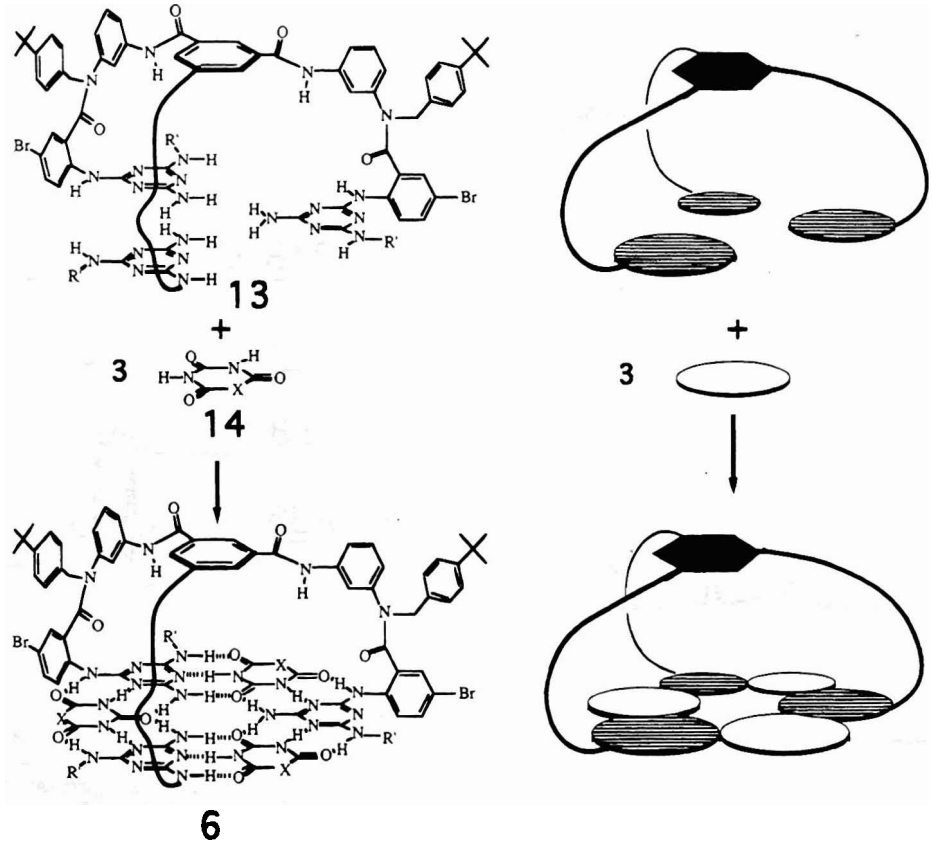


Figure 4. hub(M)$_3$ (**13**). R′ = $(CH_2)_2C(CH_3)_3$. X = NR′ (*neo*-hexylisocyanuric acid, **14**).

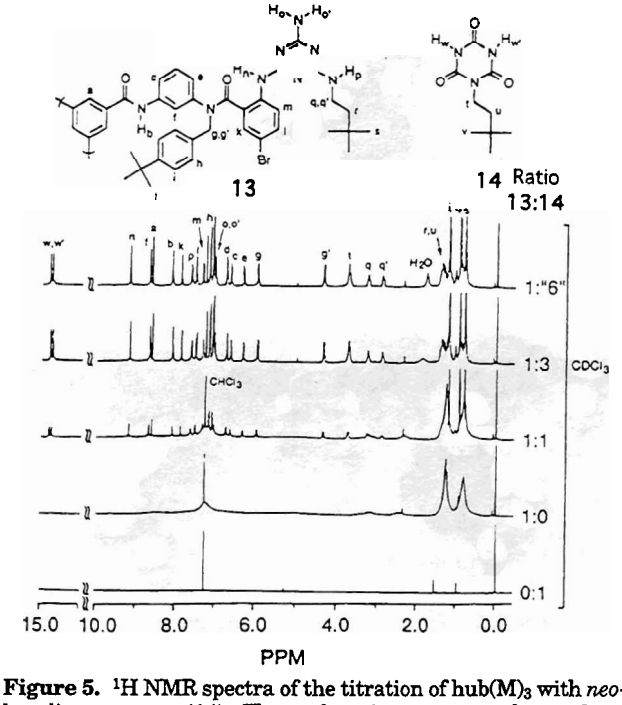


Figure 5. ^{1}H NMR spectra of the titration of hub(M)$_3$ with *neo*-hexylisocyanurate (**14**). The peak assignments are shown above the top spectrum.

Architectural Inferences. From the evidence that is now available, it is possible to draw limited but valuable generalizations describing the relationship between the structures and stabilities of the aggregates.

Geometry of Melamine-Spoke Linkage. In the aggregates based on the hub-and-spoke architecture, it increases stability to organize the uncomplexed trismelamine so that its conformation in solution in the absence of molecules of isocyanuric acid resembles that in the aggregate. Qualitatively, it is plausible that incorporation of the anthranilic acid moiety into these structures turns the melamine group more toward the center of the rosette than does a more flexible connector based on an ethylene glycol group or than does a linker in which the anthranilate is replaced by *meta*-substituted benzoic acid. Efforts to visualize the extent of this preorganization using CPK models or molecular dynamics indicate that the preorganization is not very obvious.

Analysis of the "Spokes". An interesting question for the future is the increase in stability that might be achieved if the structure of the uncomplexed hub *could* be designed to be quite close to the final aggregate. If the linker connecting the hub and the melamine groups is too flexible, the stability of the final complex is significantly decreased; there is a greater loss of conformational entropy in the flexible linker on aggregation than for a more rigid linker. The marked difference in stabilities between flex(M)$_3$ (**5**) and hub(M)$_3$ (**6**) demonstrates the importance of conformational entropy. Both molecules incorporate anthranilate linked to melamines at the rosette and to a benzene-1,3,5-tricarbonyl hub. The less stable flex(M)$_3$ utilizes propane-1,3-diol as the linker; the more stable hub(M)$_3$ utilizes a *m*-phenylenediamine. The difference between the two is three additional, freely rotating bonds in the former that must be restricted conformationally on forming the rosette. This contribution to stability from freezing three free rotors ($T\Delta S = 3RT \ln 3 = 6$ kcal/mol) is consistent with the observed differences in stability.[32]

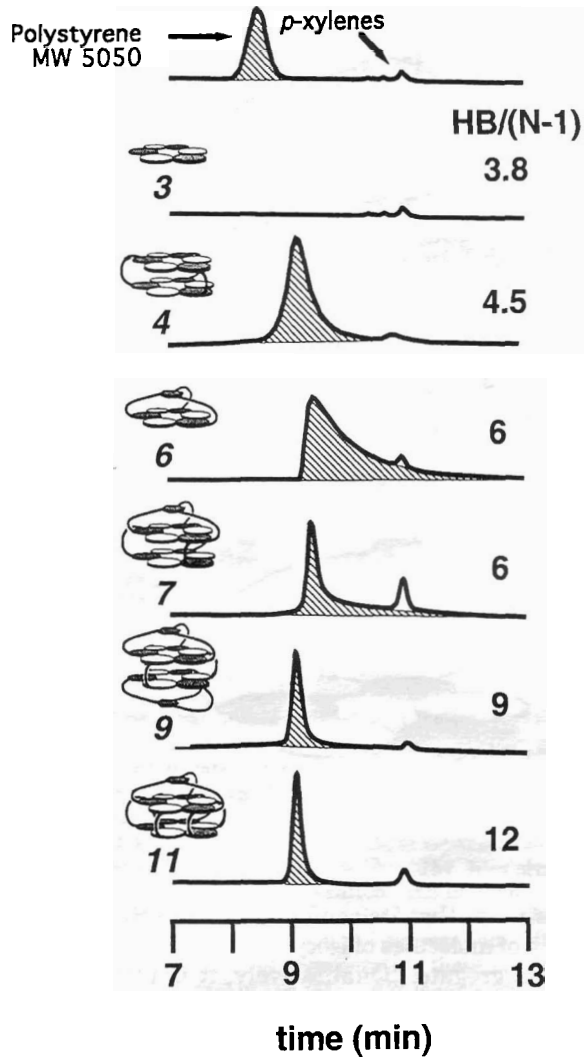


Figure 6. Comparisons of the GPC traces of various aggregates reveal information about sizes and stabilities of these aggregates. The sharp peak with little tailing suggests that **11** is the most stable. A similar retention time is indicative of similar size and shape. The unshaded peak corresponds to *p*-xylenes. Aggregate **4** shows a surprisingly sharp peak for an assembly with such a low value for HB/(N − 1). It is unclear whether **4** is anomolously stable or whether the loss of one component of the assembly results in complete dissociation and precipitation on the column.

Conformational Isomerism. Examination of the low-temperature ^{1}H NMR spectra of the complexes shows more than one conformation. Almost all of the aggregates investigated so far can be analyzed in terms of mixtures of rosettes having C_3 or C_1 symmetry in the points of attachment of the spokes to the components of the rosette (Figure 7). The ease with which the downfield region (13−16 ppm) of the proton spectrum can be analyzed, combined with the ability of this region to show the local symmetry of the hydrogen-bonded rosette, make low-temperature ^{1}H NMR spectroscopy particularly valuable as a technique for characterizing structural details of these aggregates.[31]

X-ray Structural Analysis. We have been able to obtain only one relevant structure (Figure 8).[29] This

(32) Theoretical analyses of aggregation: (a) Williams, D. H.; et al. *J. Am. Chem. Soc.* **1991**, *113*, 7020. (b) Page, M. I.; Jencks, W. P. *Gazz. Chim. Ital.* **1987**, *117*, 455. (c) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325.

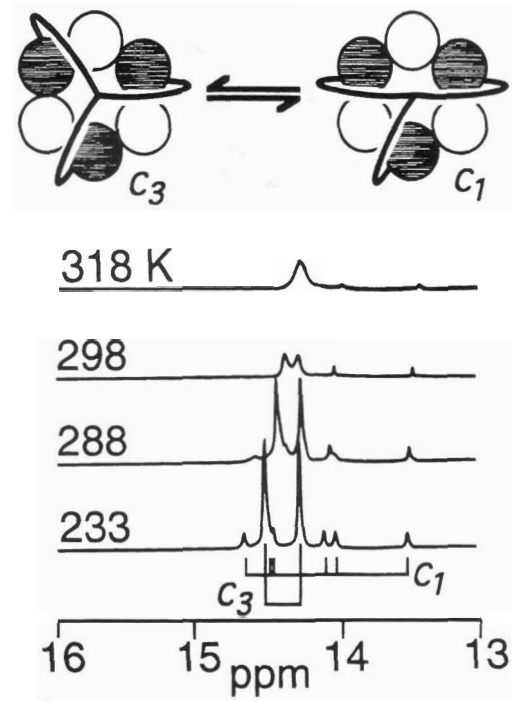


Figure 7. Low-temperature ^{1}H NMR spectroscopy resolves two conformational isomers—one having C_3 symmetry and the one having C_1 symmetry—of hub(M)$_3$:3barbital. The two isomers are represented at the top of the figure (as viewed from above). The resonances between 13 and 16 ppm are the imide resonances that are identified with numbers. The C_3 isomer has two nonequivalent imide[31] resonances of barbital while the C_1 isomer has six nonequivalent resonances.

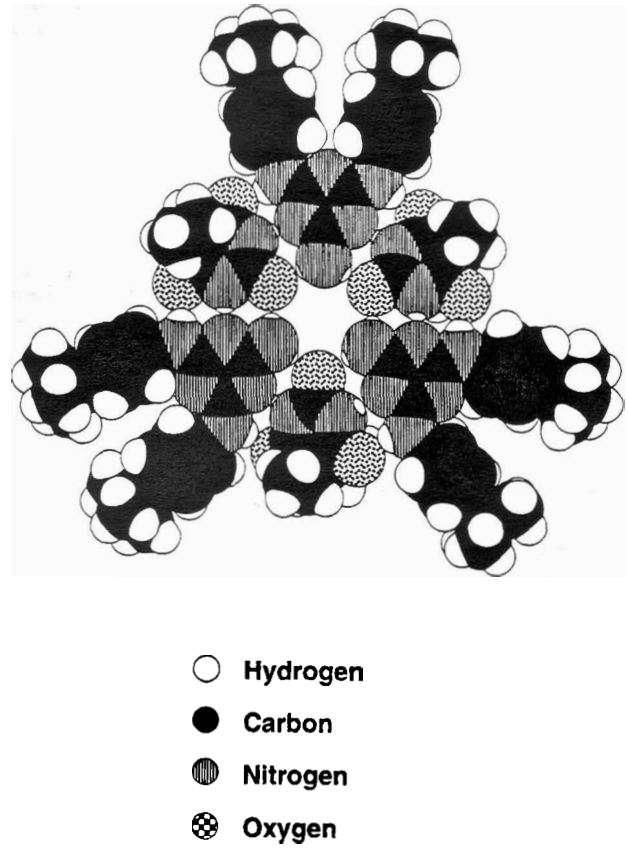


Figure 8. Crystal structure of the rosette composed of barbituric acid and bis(*p-tert*-butylphenyl)melamine.

structure establishes the basic parameters characterizing the rosette motif; it also provides an existence

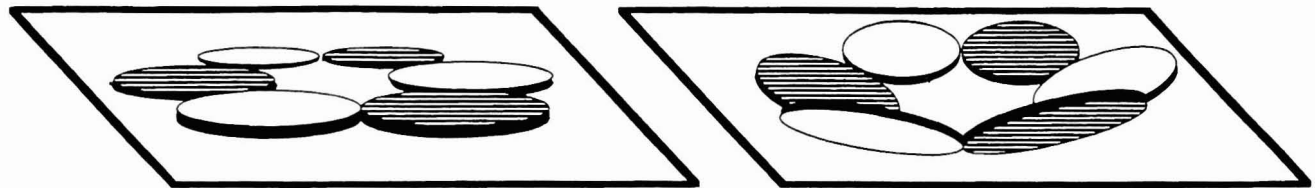


Figure 9. Differences in the planarity of the rosette of an aggregate (as calculated using molecular dynamics) can be correlated with the stability of that aggregate. An aggregate observed to have a rosette with an average structure close to planar (left) is more stable than one in which the rosette is nonplanar or fluctuates from planarity (right).

proof for an aggregate stabilized by peripheral steric interactions. Analogs of this structure prepared with smaller substituents on the para-position of the phenyl rings crystallize in linear or crinkled tape motifs.[29]

Stability. We use an empirical parameter HB/(N − 1) in making educated guesses about the probable stability of a new aggregate (Figure 3). HB is the number of hydrogen bonds in the aggregate and is related to the enthalpy of formation of the aggregate; N is the number of separate molecules present in the aggregate, and N − 1 is related to the loss in translational and rotational entropies on forming the aggregate. The detailed form of this parameter cannot be justified theoretically.[32] It has, nonetheless, a useful intuitive basis: as the number of hydrogen bonds increases, HB becomes larger and the aggregate is expected to become more stable due to the increased enthalpy of formation of hydrogen bonds (other considerations being equal); as the number of particles increases, N − 1 becomes larger, and the aggregate becomes less stable due to increased losses of translational/rotational entropies of the components on aggregation (other considerations again being equal). In general, rankings of stability expected from HB/(N − 1) agree qualitatively with those from GPC and NMR spectroscopy.

In addition, certain of the aggregates—those that we infer to be the most stable—can be dissolved in chloroform containing methanol without decomposition. For example, the ^{1}H NMR spectrum of aggregate **9** is virtually unchanged on going from pure chloroform to chloroform containing 20% (v:v) methanol; at higher concentrations of methanol, it dissociates into its components. Variable-temperature NMR spectroscopy also shows pronounced line broadening for most aggregates comprising barbiturates at temperatures above 270 K; similar spectra for isocyanurates show broadening above 300 K. Although we have not determined the origin of the broadening, it must reflect processes that interchange the environments of the N−H moieties at rates on the order of $10-10^3$ s^{-1}.

Computational Simulation. We have explored molecular dynamic simulations in an effort to develop a more quantitative basis for estimating stabilities of aggregates. These simulations are technically complicated for several reasons: the aggregates are large; solvent must be included at some level of detail; simulations cannot be run for sufficiently long times to study equilibrations involving different species directly; molecular dynamics is unreliable in generating free energies of formation for systems in which entropic terms are important. We have, however, identified a valuable qualitative *surrogate* for stability: viz., the time-averaged conformational mobility of the hydrogen-bonded rosette. We reasoned that a *stable* aggregate would be one in which a planar rosette was a clearly defined minimum in free energy and that *instability* in an aggregate might be reflected in buckling of the rosette away from planarity or in large excursions of the individual melamine or cyanuric acid components away from the mean plane of the rosette. An initial study suggests that the mean deviation from planarity of the components of the rosette correlates qualitatively with other measures of stability (Figure 9).[33]

Lessons Confirmed or Learned. Our experiments with aggregates based on the CA·M lattice have confirmed inferences about the stability of noncovalent aggregates made by others and added additional detail useful in the design of new, stable aggregates. These "lessons learned" begin to form a checklist for noncovalent synthesis.

Noncovalent aggregates based on the CA·M lattice can be stable in solution, but their design must explicity minimize the unfavorable entropy of aggregation. The most important experimental fact from this work is that these aggregates *are* stable in chloroform solution and that the CA·M lattice is a synthetically convenient platform with which to develop strategies for noncovalent synthesis. Although **8**, the largest aggregate we have prepared, has a molecular mass of only 6.4 kDa and is still far from the molecular weights of biomolecules, it incorporates 10 separate molecules! Qualitatively, aggregates having values of HB/(N − 1) > 5 seem to be stable in chloroform; those with HB/(N − 1) < 5 are only marginally stable. Aggregates with HB/(N − 1) > 9 are stable in 20% methanol/chloroform. The success of this catagorization suggests that it will eventually be possible to develop rules of thumb (and perhaps more quantitative computational tools) to estimate the relative contributions of entropy and enthalpy to stability.

Characterization of noncovalent aggregates is an exercise in strong inference rather than "proof". The most useful techniques in characterization—NMR, GPC, VPO, stoichiometry by titration—provide a convincing case for the structure of these aggregates. The case is nonetheless indirect by the rigorous standards of covalent organic chemistry, and it would be valuable to have additional techniques for characterization in solution to increase the strength of the inference of structure.

Molecular simulations are helpful in understanding these aggregates, especially their conformational entropy and solvation. Qualitative surrogates for stability may be more useful than numerical estimates. Molecular mechanics and dynamics clearly have the *potential* to be useful in understanding these ag-

(33) Chin, D. N.; Gordon, D. M.; Whitesides, G. M. *J. Am. Chem. Soc.*, in press.

gregates and especially in characterizing their molecular motions and the range of conformations open to them. It may be possible, as computers become faster, to carry our dynamics runs with sufficient precision and for sufficiently long times to obtain useful quantitative estimates of entropies; at present, it is not possible to do so, and conformational surrogates for thermodynamic parameters seem a more practical approach to theoretical investigation of stabilities than numerical estimates.

The role of solvent is more important for noncovalent assemblies than for covalent molecules. The balance between enthalpy and entropy is close in these aggregates. Small perturbations from solvation would, therefore, be expected to be much more important for them than for covalent compounds. Experimental observations indicate that addition of solvents capable of competitive hydrogen bonding completely destroys many of the aggregates. Molecular dynamics suggest a special role for solvent in these aggregates, viz., to form weak interactions that prevent the formation of patterns of hydrogen bonds within the aggregates that disrupt the rosette. The sensitivity of stability of the aggregates to solvent suggests that it may be possible to use them as probes of solvent-solute interactions.

Conclusions

Complex Aggregates Are Practical Targets for Synthesis. "Chemists make and study *molecules*". The historical strength of the field of organic chemistry is encapsulated in this statement. As organic chemistry expands to include biochemistry and materials science, we believe that *aggregates* will become increasing important targets for synthesis. The synthesis of aggregates offers practical routes to entities with high molecular weights and, perhaps, ultimately to new kinds of function, but requires new strategies for design and synthesis. Controlling noncovalent interactions and understanding the balance between entropy and enthalpy on assembly of subunits are key components of these strategies. The development of techniques for characterizing the products of noncovalent synthesis remains a challenge. Investigating the role of solvent in these systems (experimentally and computationally) will be of critical importance—especially as assemblies are designed that are intended to be soluble and stable in water.

This work was supported by NSF Grant CHE-9122331.