SELF-ASSEMBLY THROUGH NETWORKS OF HYDROGEN BONDS

J. P. Mathias, C. T. Seto, J. A. Zerkowski, and
G. M. Whitesides*

Department of Chemistry
Harvard University
Cambridge Massachusetts 02138 USA

1. INTRODUCTION

We are developing self-assembly as a strategy for the preparation of large supramolecular structures. This approach to synthesis focuses on forming networks of weak, reversible non-covalent interactions between the constituent molecules to generate a thermodynamically stable structure, rather than on forming individually strong, covalent bonds. A major intellectual impetus for the development of self-assembly in organic chemistry is the range of self-assembled structures found in living organisms.[1] Pertinent examples of these structures include (i) ^{t}RNA molecules, whose structures are dictated by series of hydrogen bonds between A:U and G:C residues,[2] and (ii) telomers, whose structures are stabilized by stacked, cyclic hydrogen bond networks between G residues.[3]

The objective of our research is the preparation of series of related self-assembled structures, using hydrogen bonds as the basic intermolecular links. We intend to use these series to address certain issues crucial to the application of self-assembly as a synthetic strategy.[4-8] These issues include:

An Understanding of Thermodynamics. Enthalpy forms the thermodynamic basis of covalent synthesis. As a consequence, the intuition of chemists concerning the structure, enthalpy, and stability is strong. The corresponding intuition regarding the relations involving entropy is generally weaker. In self-assembly, understanding the interplay between enthalpy and entropy is central to molecular design. We wish to understand this interplay in sufficient detail to predict the stability of new structures.

Considerations in Design. Preorganization of constituent molecules has emerged as a key consideration in the design

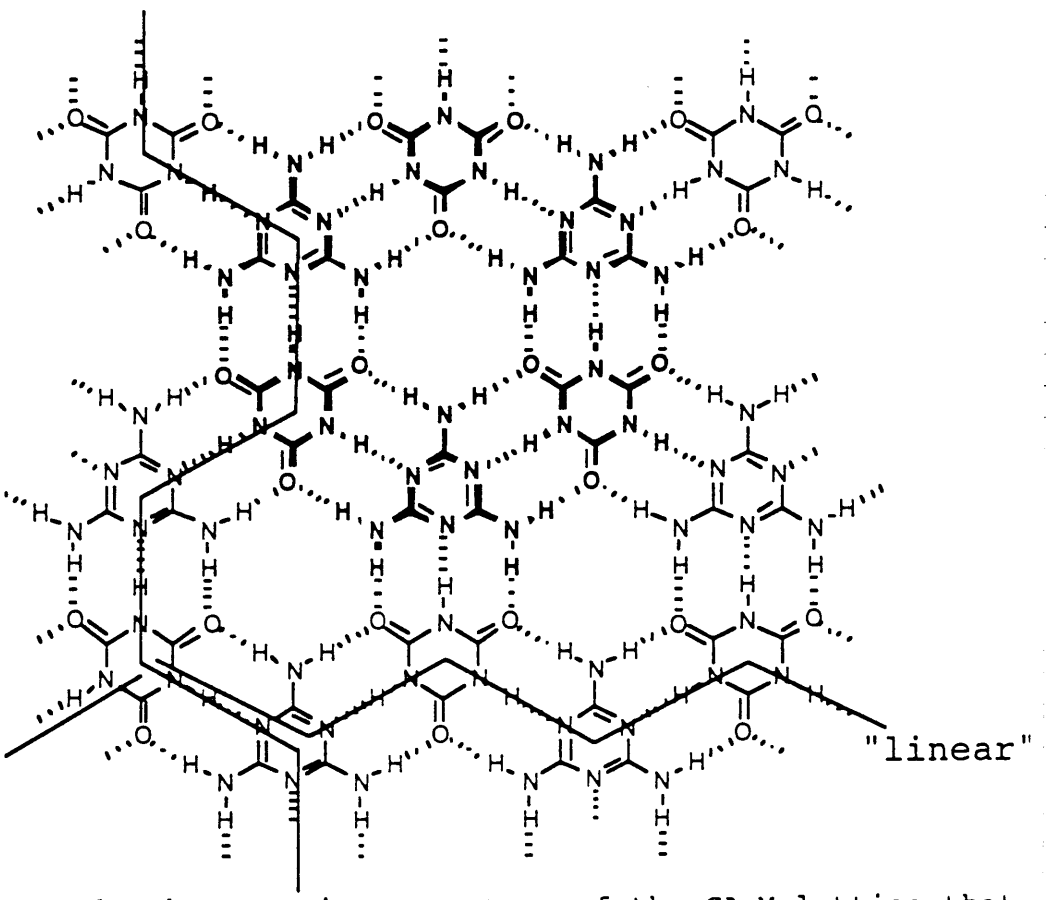


Figure 1 The putative structure of the CA·M lattice that is believed to form on mixing equimolar ratios of cyanuric acid (CA) and melamine (M). The motifs we have extracted are (i) the cyclic CA_3M_3 unit (bold), (ii) the linear tape, and (iii) the crinkled tape.

of self-assembling structures. We wish to ascertain where the balance lies between the rigidity/flexibility of constituent molecules, the ease of their synthesis, and the stabilities of complexes derived from them.

Techniques for Characterization. The characterization of large molecular structures is difficult. We wish to determine which analytical methods provide the most useful data for the characterization of non-covalently bound supramolecular aggregates in organic solution.

2. SOLUBLE HYDROGEN-BONDED COMPLEXES BASED ON THE CA·M LATTICE

The strategy we have adopted for the formation of hydrogen-bonded supramolecular aggregates in organic solution is based on the synthesis of discrete, soluble portions of the hydrogen-bonded lattice between cyanuric acid and melamine (CA·M) (Figure 1).[9] The motif we have

used to build self-assembling structures is the cyclic CA_3M_3 unit (bold) that characterizes this lattice. The CA·M lattice fulfills two key requirements to serve as the basis of self-assembling structures. First, the high density of hydrogen bonds in CA·M provides a strong enthalpic driving force for self-assembly. Second, the cyclic CA_3M_3 unit is a highly symmetrical motif. This factor aids characterization of these supramolecular aggregates, especially by n.m.r. spectroscopy.

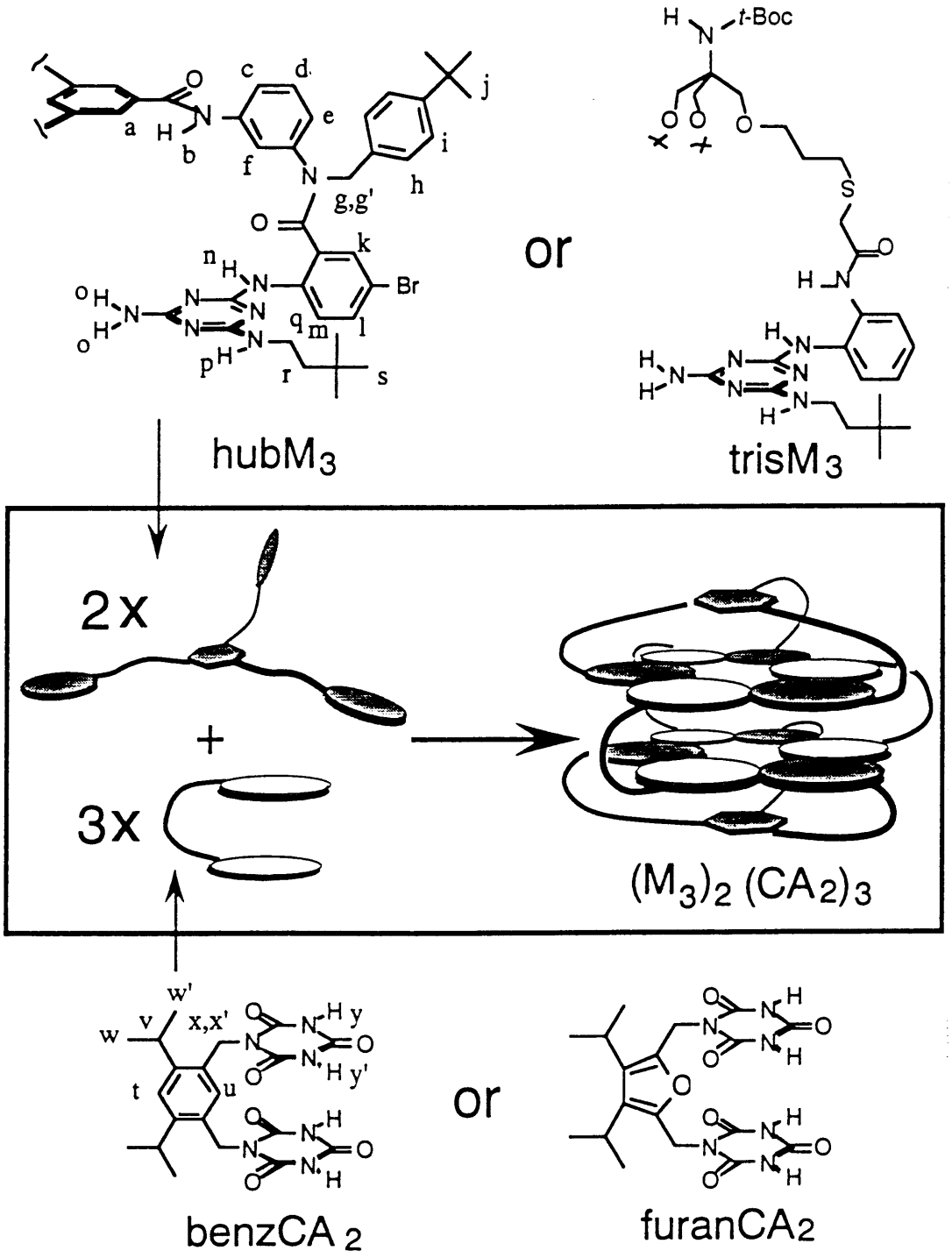
Preparation of 2+3 Supramolecular Aggregates.

Mixing equimolar portions of isocyanurates and melamines in chloroform at room temperature results in their association to form random oligomers, with no evidence for the existence of a stable cyclic CA_3M_3 unit in solution. To reduce the unfavorable translational entropy that is associated with self-assembly, and to reduce the loss of conformational entropy upon complexation (by attempting to impose the bound conformation on the unbound constituents), we have used a 'spoke' and 'hub' architecture to preorganize both melamine and isocyanurate units prior to binding.[10] This approach to self-assembly is illustrated in Scheme 1. Two equivalents of the trivalent melamine derivatives, $hubM_3$ and its more flexible analog $trisM_3$, self-assemble with three equivalents of the dimeric isocyanurate derivatives, $benzCA_2$ and $furanCA_2$, to afford 2+3 adducts in chloroform solution.[11] These complexes are stabilized by 36 hydrogen bonds between the five constituent molecules. We have used five analytical techniques to characterize these 2+3 supramolecular aggregates.

1. Solubility: The solubility of the bisisocyanurate derivatives in chloroform is low (<0.1 mM). On complexation with $hubM_3$, the bisisocyanurates dissolve. The ability of two equivalents of $hubM_3$ to solubilize three equivalents of $benzCA_2$ provides qualitative evidence for the stoichiometry of the resulting complex.

2. UV Spectroscopy: The changes in λmax that occur on addition of $benzCA_2$ to $hubM_3$, up to the stoichiometry required for formation of the 2+3 complex, provide quantitative evidence for the stoichiometry of the complex. Beyond this point, no further changes occur.

3. ^1H N.M.R. Spectroscopy: The solubility of these complexes in chloroform has allowed their structural characterization by ^1H n.m.r. spectroscopy. Figure 2 shows the changes that occur in the ^1H n.m.r spectrum in chloroform on going from uncomplexed $hubM_3$ to $(hubM_3)_2(benzCA_2)_3$. The spectrum of $hubM_3$ is broad as a result of self-association and of hindered rotation about



Scheme 1 Self-assembly of the trivalent melamine derivatives, $hubM_3$ and $trisM_3$, with the bivalent isocyanurate derivatives, $benzCA_2$ and $furanCA_2$, to form 2+3 supramolecular aggregates. The molecular weight of $(hubM_3)_2(benzCA_2)_3$ is 5.519 KDa and the molecular weight of $(trisM_3)_2(benzCA_2)_3$ is 4.266 KDa.

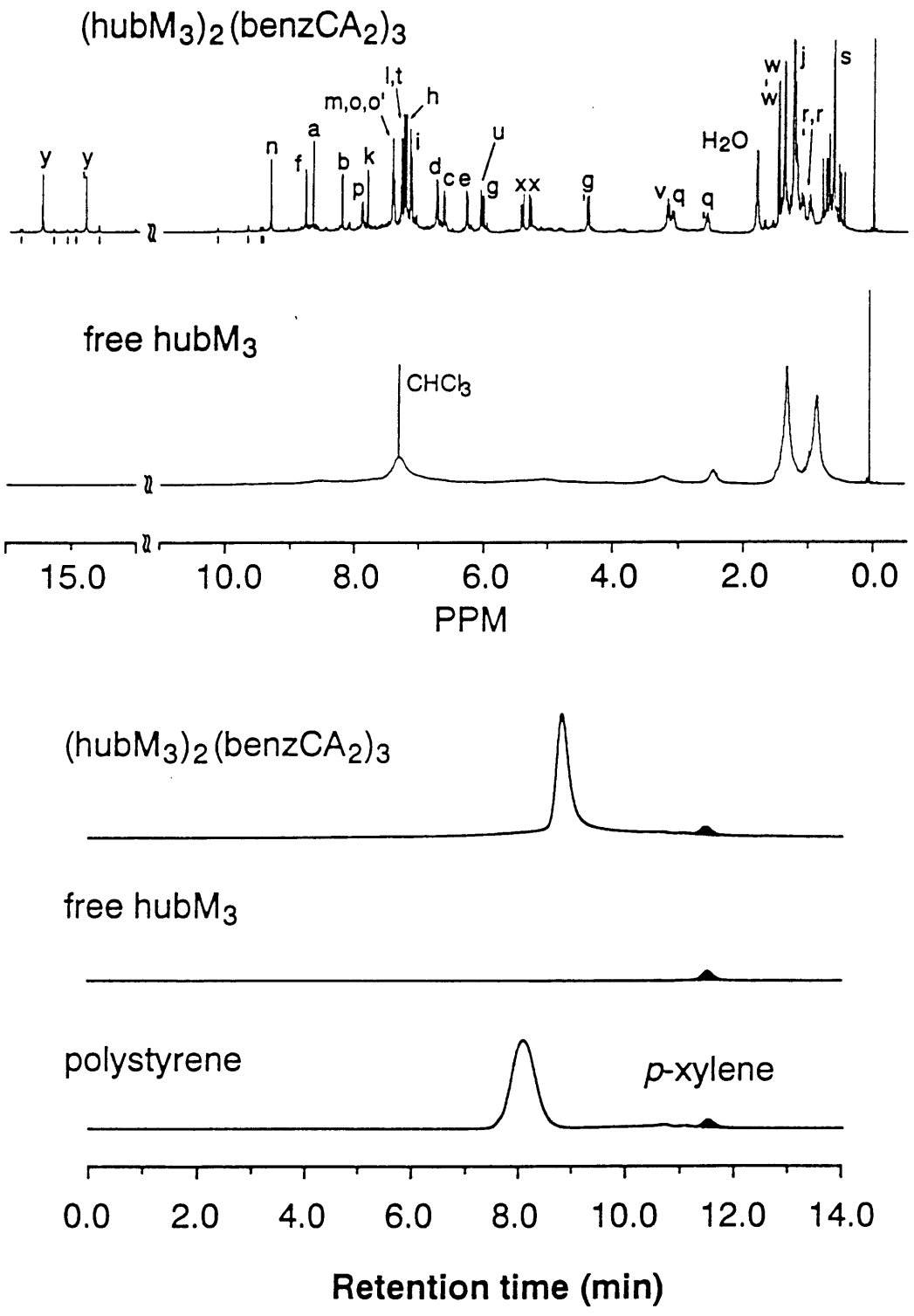


Figure 2 Characterization of $(hubM_3)_2(benzCA_2)_3$ by ^1H n.m.r spectroscopy and gel permeation chromatography. The annotation on the top ^1H n.m.r. spectrum corresponds with that for $(hubM_3)_2(benzCA_2)_3$ shown in Scheme 1. The retention time in g.p.c. increases as the size of the molecle decreases. The polystyrene standard has a mean molecular weight of 5050 and a polydispersity of 1.05.

the amide and triazine-NH bonds. In contrast, the resonances in the spectrum of $(hubM_3)_2(benzCA_2)_3$ are sharp and predominantly (>95%) those of a single species. These resonances are consistent with the proposed structure of $(hubM_3)_2(benzCA_2)_3$. At intermediate stoichiometries, the spectra show only resonances associated with $(hubM_3)_2(benzCA_2)_3$ and uncomplexed $hubM_3$. This observation is consistent with the operation of cooperativity in the formation of $(hubM_3)_2(benzCA_2)_3$. In contrast to $(hubM_3)_2(benzCA_2)_3$, the more flexible $(trisM_3)_2(benzCA_2)_3$ aggregate exhibits considerable conformational isomerism. Full structural assignment of this complex is, therefore, impossible. This feature illustrates the importance of preparing supramolecular aggregates that are highly structured and symmetrical.

4. Gel Permeation Chromatography (g.p.c.): G.p.c. is a technique that separates molecules on the basis of their hydrodynamic radii. It is a useful technique for analyzing non-covalently bound structures in organic solution, since it provides information (albeit qualitative information) about both stability and size. Figure 2 shows the g.p.c. traces of uncomplexed $hubM_3$, $(hubM_3)_2(benzCA_2)_3$, and a polystyrene standard in chloroform. Uncomplexed $hubM_3$ exhibits self-association in solution and exists as a poorly-defined mixture of aggregates. The g.p.c. trace of $hubM_3$ is very broad and featureless. In contrast, the g.p.c trace of $(hubM_3)_2(benzCA_2)_3$ shows a single, sharp peak that is consistent with the presence of a single species in solution. The profile of the peak suggests that the stability of the complex is high, with no evidence of decomplexation over the duration of the run. We believe that the shapes of peaks for different hydrogen bonded aggregates in g.p.c. provide a useful qualitative measure of their relative stabilities.

5. Vapor Pressure Osmometry (v.p.o.): V.p.o. is a technique for obtaining molecular weights of molecules in solution, by correlating the vapor pressure of a solution containing a compound of unknown molecular weight to that of solutions of standards of known molecular weights. We have determined the molecular weights of our 2+3 supramolecular aggregates in chloroform solution, to an accuracy of ± 10%, by correlating results against those of N,N′ bis^{t}Boc-gramicidin-S (MW 1342), sucrose octaacetate (MW 679), polystyrene (MW 5050), and perbenzoyl β-cyclodextrin (MW 3321). We observe a significant range of values for molecular weights depending on the standard used. We are investigating the effects that self association and solvent association, caused by hydrogen bond sites on the periphery of our molecules, have on the structures in solution (and, therefore, on the inferred molecular weights) of these complexes. An improved

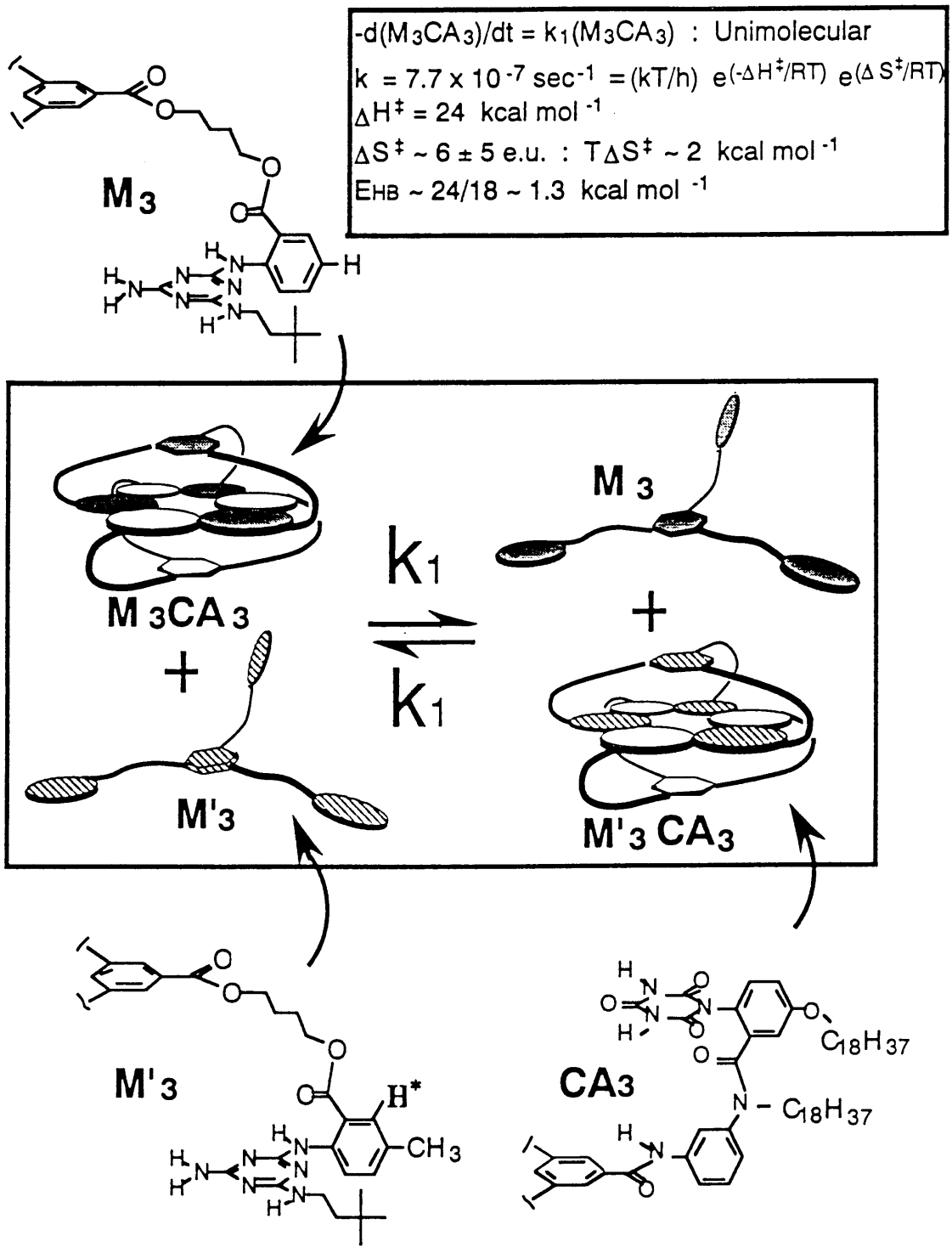
understanding of these features should allow us to apply v.p.o. to the analysis of supramolecular aggregates with greater accuracy.

An Assessment of the Thermodynamic Parameters of Self-Assembly.

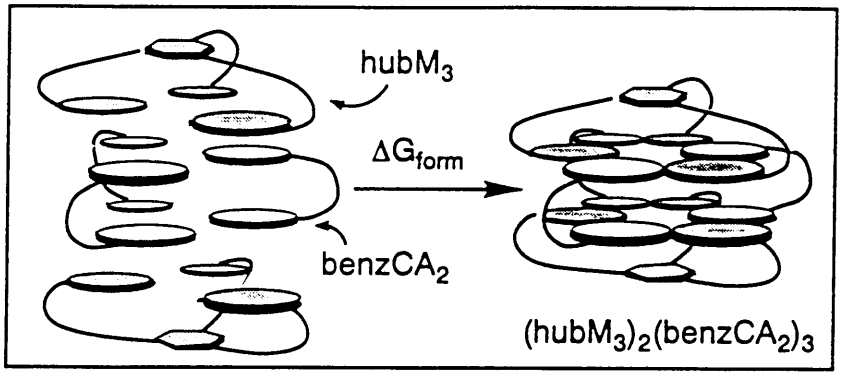
To improve our understanding of the thermodynamics of self-assembly, we have examined the rates of exchange of components in self-assembled complexes with analogous structures, uncomplexed in solution, by ^{1}H n.m.r spectroscopy (Scheme 2). The 1+1 complexes formed between the trivalent melamine derivatives M_3 and the complementary trivalent isocyanurate derivatives CA_3 are the most stable we have prepared. By measuring the rate constant for exchange at various temperatures, we can obtain the thermodynamic parameters that are associated with the transition state of the exchange process. These values are given in Scheme 2. The exchange process is unimolecular. This observation suggests that the transition state for exchange corresponds to the fully decomplexed state. Dividing the enthalpy of activation for this exchange (*i.e.* for the dissociation reaction) (24 kcal mol^{-1}) by the number of hydrogen bonds (18) gives an individual hydrogen bond enthalpy of 1.3 kcal mol^{-1}. Although this analysis contains some serious and debatable assumptions, the value of the hydrogen bond strength compares favorably with values obtained by other authors.[12-14]

Qualitative Predictions of Stability.

Based on the enthalpy for each hydrogen bond from the previous section, we have constructed a semi-quantitative analysis of the relative contributions of different free energy terms to self-assembly processes. An example, based on the assembly of $(hubM_3)_2(benzCA_2)_3$, is shown in Scheme 3. The enthalpy of formation of 36 hydrogen bonds in the two cyclic CA_3M_3 units should be -48 kcal mol^{-1}. The term for translational entropy represents the loss in entropy on bringing the five molecules together in space. The term for rotational entropy governs the loss in entropy on bringing together the binding regions of the melamine and isocyanurate pieces. The term for conformational entropy relates to the loss in entropy on restricting rotation about the flexible bonds in $hubM_3$. We estimate that each arm of $hubM_3$ contains 7 bonds that must be frozen into one of two rotamers. The change in conformational entropy on constraining two $hubM_3$ molecules is, therefore, $2 \times (21 \times -RT\ln(1/2)) = -17$ kcal mol^{-1}. $BenzCA_2$ has no rotamers that are restricted as a consequence of binding. Combining these values leads to a predicted free energy of formation of ~ -13 kcal mol^{-1} for $(hubM_3)_2(benzCA_2)_3$.



Scheme 2 The thermodynamic parameters associated with the transition state for the exchange process illustrated. The reaction was followed by monitoring the signal of the proton indicated on M'$_3$ by an asterisk by ^{1}H n.m.r. spectroscopy.



Enthalpy of Formation : ΔH_{formn} ~ -48 kcal mol^{-1}

Translational Entropy : $-T\Delta S_{trans}$ ~ +12 kcal mol^{-1}

Rotational Entropy : $-T\Delta S_{rotn}$ ~ + 6 kcal mol^{-1}

Conformational Entropy : $-T\Delta S_{confn}$ ~ +17 kcal mol^{-1}

ΔG_{form} ~ -48 + 35 ~ **-13 kcal mol^{-1}**

Scheme 3 Semiquantitative estimates of the contributions of enthalpy and entropy to the free energy of self-assembly for $(hubM_3)_2(benzCA_2)_3$.

A second qualitative predictor that we have applied to our systems is the number HB/(N-1). HB is the number of hydrogen bonds in the aggregate. N is the number of constituent molecules in the aggregate. The larger the value of HB, the more stable the aggregate; the larger the value of N, the less stable the aggregate. The ratio HB/(N-1) is not correlated in a fundamental way to ΔG, but HB correlates with the enthalpy of formation of the network of hydrogen bonds, and (N-1) correlates with the loss of translational entropy on formation of the aggregate. Large values of HB/(N-1) should, therefore, correlate with high stability for these types of complexes. Figure 3 shows the values of HB/(N-1) for six types of our supramolecular aggregates. These structures range from the untethered cyclic CA_3M_3 unit to the very stable 1+1 complex. The trend in stabilities that is suggested in this figure matches that we have observed.

Although these methods for assessing the thermodynamic aspects of self assembly are both semi-quantitative, they do represent useful methods of

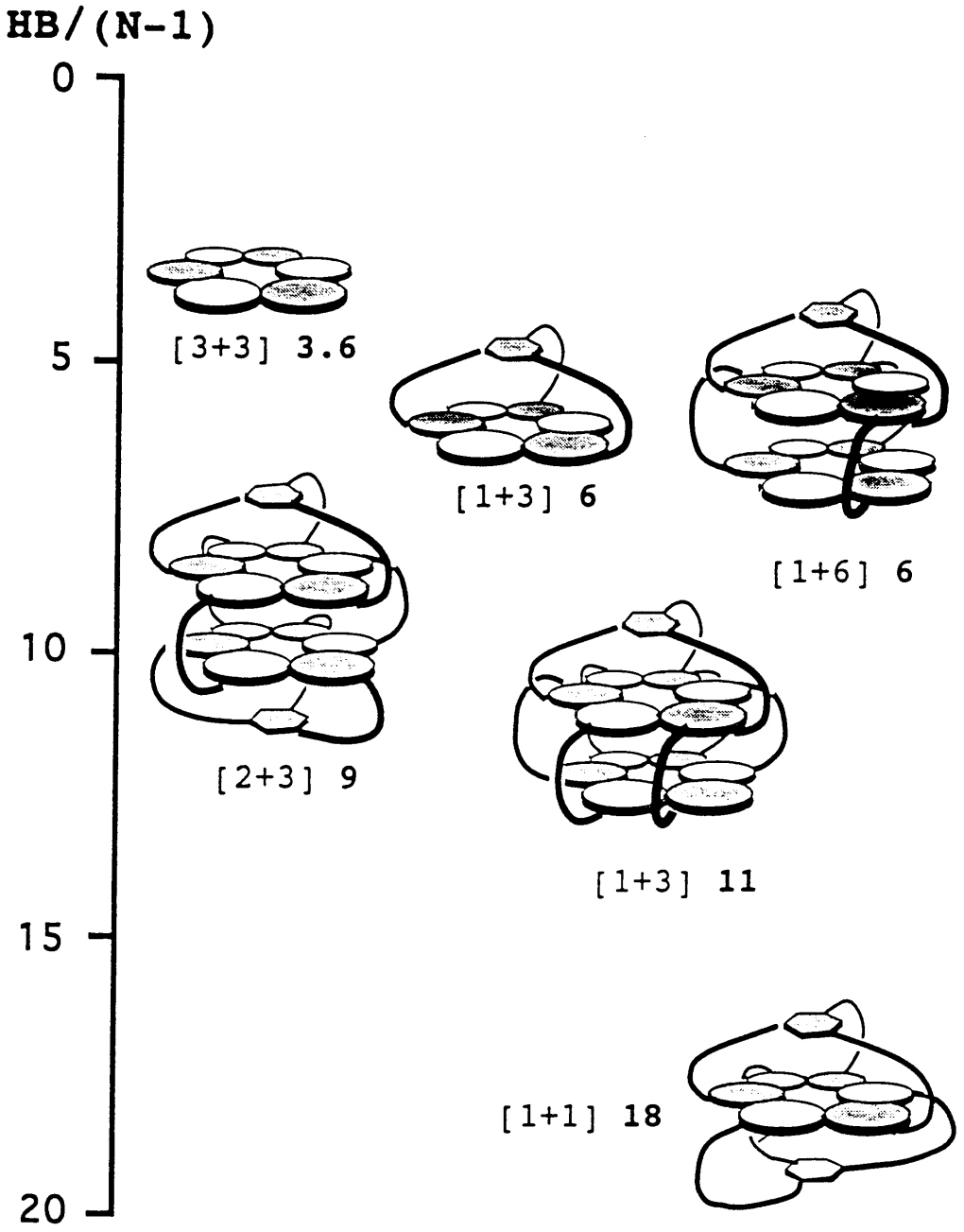


Figure 3 Values of HB/(N-1) for six self-assembled supramolecular aggregates. An increase in the value of HB/(N-1) should indicate in increase in the stability of the supramolecular aggregate. The order of these values of HB/(N-1) correlates qualitatively with the stabilities of the complexes. The numbers in brackets refer to the numbers of melamine units (first) and isocyanurate units (second) involved in the supramolecular aggregate.

predicting which systems may self assemble to give stable structures and which systems may have low stability.

3. CO-CRYSTALLIZATION : HYDROGEN-BONDED SELF-ASSEMBLY IN THE SOLID STATE.

The vast literature on hydrogen bonding in crystals is a testament to the power of this interaction as a stabilizing feature in the solid state.[15,16] Crystals that contain extensive networks of hydrogen bonds provide a means to investigate the effects of molecular structure on solid-state packing, and, ultimately, to design and build useful solids based on molecular crystals. This strategy for crystal engineering requires a step beyond the level of molecular design used to prepare soluble molecules, and must be able to tackle problems that are unique to the solid state, such as crystal packing forces.

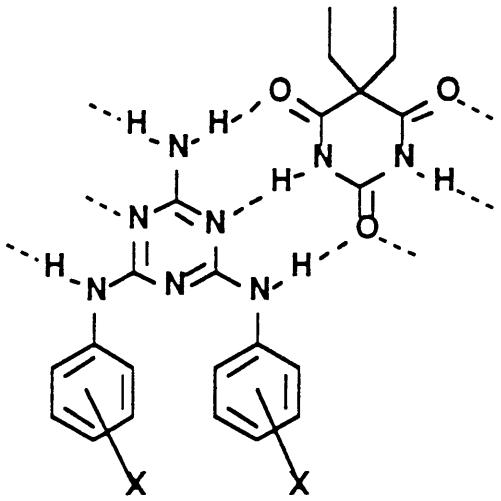
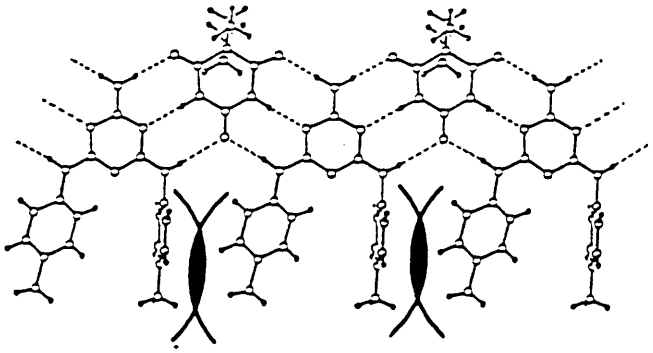


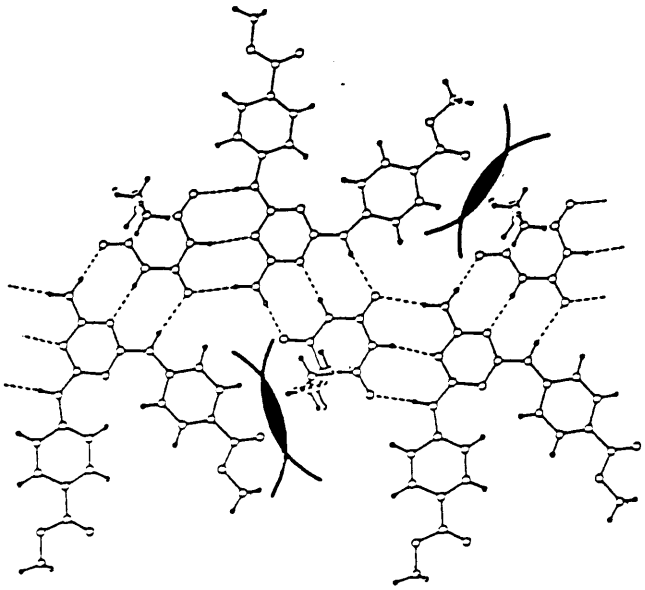
Figure 4 The series of co-crystals we have obtained demonstrates the effect of changing the phenyl substituent (X) on the crystal structure of the 1:1 complex between the substituted diphenylmelamine derivatives and barbital. In all cases, the triad hydrogen bond network between the substituted diphenylmelamines and barbital is retained.

The basis for our efforts in crystal engineering is the collection of motifs found within the CA·M lattice (Figure 1). By using substituted melamines and isocyanurates, we can interfere with the hydrogen bond network and 'excise' portions of the lattice. We have used series of co-crystals formed from equimolar ratios of substituted diarylmelamine derivatives and barbital (Figure 4) to identify three motifs, (i) linear tapes, (ii) crinkled tapes, and (iii) rosettes based upon the cyclic CA_3M_3 structure.[9] These motifs are shown in Figure 5. They represent a basis set that we are using to probe systematically the interactions that determine the structure of the crystal. We have shown that small changes in the size and position of the substituent on the phenyl ring of the diarylmelamine derivatives can be performed without disrupting the triad pattern of hydrogen

'Linear Tape'
X = *para*-Me



'Crinkled Tape'
X = *para*-CO_2Me



'Rosette'
X = *para*-tBu

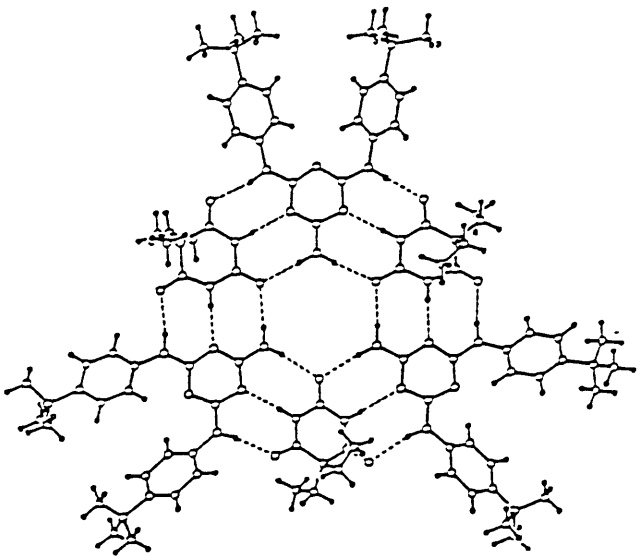


Figure 5 The effects on the crystal structure of the steric requirements of the *para*-substituents on the diarylmelamine derivatives. We propose that the differentiation of linear tape, crinkled tape, and rosette motifs is determined by the unfavorable steric interactions that are indicated on the structures.

bonds between adjacent pieces.[17] Instead, this substitution influences the motif adopted in the crystal. Co-crystals barbital and *para*-(F, Cl, Br, Me, CF_3) and *meta*-(F, I, Me, CF_3) substituted diarylmelamines adopt the linear tape motif; co-crystals between *para*-CO_2Me-diarylmelamine and barbital adopt the crinkled tape motif; co-crystals between *para*-^tBu-diarylmelamines and barbital adopt the rosette motif. We have used steric arguments to rationalize the interactions that are responsible for the differentiation between the respective motifs. We consider the linear tape to be the lowest energy motif based on its prevalence. The crinkled tape is adopted when steric interactions between adjacent *para*-substituents on the diarylmelamines become unfavorable in the linear tape. These interactions are relieved in the crinkled tape. Further increases in the steric bulk of the *para*-substituents create an unfavorable contact with the proximal ethyl groups of the barbital moieties. In this case, the steric strain is relieved by the adoption of the rosette motif, in which the substituents on the diarylmelamine groups are at their most distant.

4. CONCLUSIONS AND FUTURE PROSPECTS

We have prepared a series of supramolecular aggregates based on networks of hydrogen bonds. These structures demonstrate the versatility of the CA·M lattice as the basis for constructing self-assembling supramolecules. In the solution state, these molecules are characterized by 18 or 36 hydrogen bonds, 4-7 constituent pieces, molecular weights in the range 3-7 KDa, various levels of preorganization, and a range of stabilities. We are now in a position to (i) test certain hypotheses (for example the relationship of HB/(N-1) to stability); (ii) assess the relative stabilities of many of these complexes by ^1H n.m.r. exchange experiments; (iii) assess critically our criteria for design (in particular, how effective is our preorganization?); (iv) apply molecular modeling to assist our design and to correlate structural information; (v) prepare structures with increasing stabilities, and consequently; (vi) prepare structures of increasing size. Most importantly, we have obtained a series of closely-related self-assembled structures that is allowing us to begin to decipher trends and make predictions about new supramolecular aggregates.

In the solid state, we have applied an approach to crystal design based upon a closely-related series of structures. The central features of this program are (i) the observation of trends and (ii) the development of a capability for prediction. We believe that we can predict secondary architecture--for example, the occurrence of tapes versus rosettes (Figure 5)--with some confidence, based on steric arguments . We are now tackling two important issues. The first problem is that

of polymorphism, *i.e.* how many crystal phases are thermodynamically accessible? Which are kinetically preferred? The use of powder diffraction methods and crystallization from different solvents should determine whether the crystal structures we have observed are determined by kinetic or thermodynamic factors. The second issue is the control of tertiary architecture. Can we prepare designed solids by introducing new levels of orthogonal non-covalent interactions that will control the crystal packing of our basic tape motifs?

Molecular self-assembly is in its infancy. Any real application of self-assembly, in either solution or solid states, will depend on the availability of fundamental information concerning thermodynamics, stabilities of complexes, rules for design, and techniques for characterization. When this fundamental information is available, it will be possible to design and synthesize new types of structures efficiently.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (Grants CHE-88-12709 to GMW, DMR 89-20490 to the Harvard University Materials Research Laboratory, and CHE-80-00670 for the purchase of a Siemens X-Ray Diffractometer). NMR instrumentation was supported by the National Science Foundation Grant CHE-84-10774. The Harvard University Chemistry Department Mass Spectrometry Facility was supported by National Science Foundation Grant CHE-90-20043 and National Institute of Health Grant 1 S10 RR06716-01. JPM acknowledges support from an SERC/NATO Postdoctoral Fellowship, 1991-93. CTS was an Eli Lilly Predoctoral Fellow, 1991. We thank Professor Robert Cohen (MIT, Chemical Engineering) for loan of the vapor pressure osmometer.

REFERENCES

1. J.S. Lindsey, New J. Chem, 1991, 15, 153.
2. D.E. Draper, Acc. Chem. Res., 1992, 25, 201.
3. C. Kang, X. Zhang, R. Ratliff, R. Moyzis, and A. Rich, Nature, 1992, 350, 126.
4. G.M. Whitesides, J.P. Mathias, and C.T. Seto, Science, 1991, 254, 1312.
5. J.-M. Lehn, Angew. Chem. Int. Ed. Engl., 1990, 29, 1304.
6. P. Tecilla, R.P. Dixon, G. Slobodkin, D.S. Alavi, D.H. Waldbeck, and A.D. Hamilton et al., J. Am. Chem. Soc., 1990, 112, 9408.
7. J.F. Stoddart et al., J. Am. Chem. Soc., 1992, 114, 193.
8. S.C. Zimmerman and B.F. Duerr, J. Org. Chem., 1992, 57, 2215.

9. J.A. Zerkowski, C.T. Seto, and G.M. Whitesides, J. Am. Chem. Soc., 1992, In press.
10. C.T. Seto, and G.M. Whitesides, J. Am. Chem. Soc., 1990, 112, 6409.
11. C.T. Seto, and G.M. Whitesides, J. Am. Chem. Soc., 1991, 113, 712.
12. (a) L.D. Williams, B. Chaula, and B.R. Shaw, Biopolymers, 1987, 26, 591.
13. A.R. Fersht, Trends in Biochem. Sci., 1987, 12, 301.
14. H.-J. Schneider, Angew. Chem. Int. Ed. Engl., 1991, 30, 1417.
15. C.R. Desiraju, 'Crystal Engineering: The Design of Organic Solids', Elsevier, Amsterdam, 1989.
16. M.C. Etter, J. Phys. Chem., 1991, 95, 460.
17. J.A. Zerkowski, C.T. Seto, D.A. Wierda, and G.M. Whitesides, J. Am. Chem. Soc., 1990, 112, 9025.