

A Model for Self-Assembly in Solution¹

Gianfranco Ercolani[†]

Dipartimento di Scienze e Tecnologie Chimiche, Università di Roma Tor Vergata,
Via della Ricerca Scientifica, 00133 Roma, Italy

Received: December 27, 2002; In Final Form: March 12, 2003

A model treating the competition, under thermodynamic control, between self-assembly and nonlinear random polymerization is presented. The fundamental quantities on which the treatment is based are the effective molarity (EM) of the assembly and the equilibrium constant for the intermolecular model reaction between monofunctional reactants (K_{inter}). Knowledge of these quantities allows the evaluation of the distribution curve of the self-assembling complex. In order for effective self-assembly to take place, the product $K_{\text{inter}}\text{EM}$ must be no lower than a limit value easily computable on the basis of simple structural parameters such as the number of molecules in the assembly (N), the number of bonds joining them (B), and the number of interaction sites in the monomers. This limit decreases on decreasing N and on increasing B , and the most obvious way to realize this condition is by increasing the degree of cyclicality of the assembly ($B - N + 1$). The yield of an assembly with a high degree of cyclicality is very sensitive also to modest changes of $K_{\text{inter}}\text{EM}$ about its limit value. Depending on the value of the monomers concentration, the assembly could undergo either sharp disassembly (*denaturation*) or conversion into gel.

Introduction

Self-assembly consists of the spontaneous generation of a well-defined, discrete supramolecular architecture from a given set of components under thermodynamic equilibration.² A variety of molecular architectures have been obtained in this way, spanning from exotic structures such as ladders, helicates, and grids to two-dimensional or three-dimensional structures that closely resemble well-known geometric shapes, including triangles, squares, hexagons, cubes, triangular prisms, octahedra, cuboctahedra, etc. Such materials are interesting as artificial, molecular-scale containers or receptors, in which novel synthetic chemistry, electrochemistry, photoluminescent chemistry, supramolecular chemistry, or catalytic chemistry, inter alia, can be carried out.

Despite the plethora of self-assembling systems that can be found in the literature, the physicochemical basis of self-assembly is, however, largely unexplored. As a contribution to fill this gap, here is presented a model of general applicability that addresses a number of issues that are crucial to the understanding of self-assembly, inter alia: What are the factors governing self-assembly and how can they be put on a quantitative scale? What is the range of reactant concentrations in which a self-assembling structure is stable? What are the criteria to judge if a given assembly is more or less favored than another? It is apparent that the answers to these and other related questions are of fundamental importance not only to allow the deliberate formation of desired architectures but also to extract information from a system under study in a consistent and significant way.

Theoretical Basis

Self-assembly of a discrete supramolecular structure always occurs in competition with the process of polymerization. The

self-assembly of a two-dimensional structure requires bifunctional monomers that react to yield a specific cyclic oligomer in competition with linear polymerization; treatment of this case has been previously reported and experimentally tested.³ The case regarding the self-assembly of a three-dimensional structure is more complex because at least one of the reactants carries more than two reactive groups and thus the self-assembling architecture is formed in competition with nonlinear oligomers displaying an enormous number of different topologies. Classification of the various species that could form in solution is a hopeless task; however, as will be illustrated in the following, such a level of detail is not needed to extract the necessary information.

In the vast majority of the cases self-assembly of a three-dimensional architecture requires two monomeric building blocks. Let us indicate the two building blocks as L_1 , having l binding sites $-A$, and M_1 , having m binding sites $-B$. The two functional groups $-A$ and $-B$ are each capable of reacting with the other only in a reversible addition reaction. When L_1 and M_1 are mixed and the equilibrium is attained, the monomeric units can be considered as partitioned in two fractions in equilibrium between them, one constituted by an infinite number of oligomers having one or more loops in their structure and including the assembly S , and the other constituted by an infinite number of more or less branched oligomers devoid of loops. Self-assembly takes place when the monomers have a rigid structure predisposed⁴ in such a way that formation of the self-assembling complex is strongly favored over other cyclic or polycyclic species. The latter are disfavored by one or more of the following: (i) the presence of unreacted end groups, (ii) strained loops involving a high enthalpy content, and (iii) large loops involving a high entropy loss. It is assumed therefore that S is the only significant species of the first fraction. This assumption, substantiated later in this paper, is further justified by the fact that the most important theory of nonlinear

[†] E-mail: ercolani@uniroma2.it.

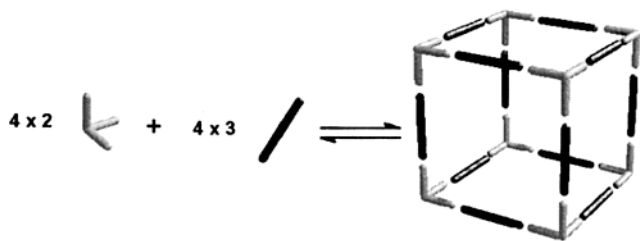


Figure 1. Self-assembly of a cube by two predisposed building blocks. In the example shown $p = 4$, $m = 2$, $l = 3$.

polymerization, due to the contributions of Flory and Stockmayer, completely neglects the formation of looped structures.⁵

The acyclic oligomers of the second fraction have the general formula L_rM_s . They can be subdivided in families of oligomers, R_i , where i , the degree of polymerization, is given by $i = r + s$, so that R_1 represents the monomers L_1 and M_1 , R_2 the dimers, and so on. Schematically the equilibria to consider for the process of self-assembly are shown in Scheme 1.

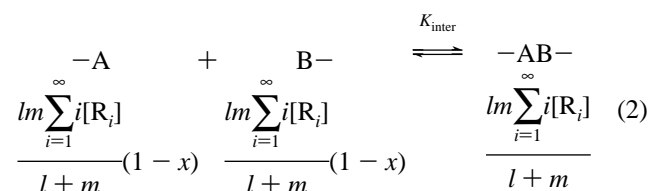
The molecular formula of the assembly S is $L_{pm}M_{pl}$, where p is a coefficient needed to account for the stoichiometry of L and M within S . A cartoon showing the self-assembly of a cube is reported, by way of illustration, in Figure 1.

The mass balance equation in terms of the total number of monomer units is then given by

$$[L_1]_0 + [M_1]_0 = (pl + pm)[S] + \sum_{i=1}^{\infty} i[R_i] \quad (1)$$

where the subscript 0 has the meaning of initial concentration.

It is obvious that the maximum concentration of S is obtained when the initial concentration of functional groups $-A$ and $-B$ is equal, i.e., when the molar fraction of the monomers, L_1 and M_1 , is equal to $m/(l+m)$ and $l/(l+m)$, respectively; therefore, only this case will be treated here. The total number of bound and unbound monomeric units L_1 in the acyclic fraction then will be $m \sum_{i=1}^{\infty} i[R_i]/(l+m)$. Because each L_1 unit carries l functional groups $-A$, the total number of functional groups $-A$ (and $-B$) in the acyclic fraction is $lm \sum_{i=1}^{\infty} i[R_i]/(l+m)$. Assuming that all the intermolecular reactions between $-A$ and $-B$ groups in the acyclic fraction occur with the same probability, measured by the equilibrium constant K_{inter} , the overall equilibrium of end groups in the acyclic fraction is expressed by eq 2, where x is the fraction of reacted functional groups in the polymer.



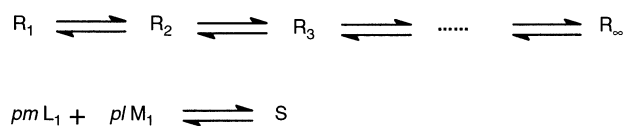
From the equilibrium in eq 2, eq 3 is obtained.

$$\sum_{i=1}^{\infty} i[R_i] = \frac{(l+m)}{lm} \frac{1}{K_{\text{inter}}} \frac{x}{(1-x)^2} \quad (3)$$

Substituting eq 3 into eq 1, and considering that the number of molecules in the assembly, N , is equal to $pl + pm$, and the number of bonds joining them, B , is equal to plm , eq 4 is obtained

$$[L_1]_0 + [M_1]_0 = N[S] + \frac{N}{B} \frac{1}{K_{\text{inter}}} \frac{x}{(1-x)^2} \quad (4)$$

SCHEME 1



Now consider the infinite McLaurin expansion shown in eq 5

$$\frac{x}{(1-x)^2} = \sum_{i=1}^{\infty} ix^i \quad (5)$$

It is evident by substituting eq 5 into eq 3, that eq 6 holds.

$$[R_i] = \frac{(l+m)}{lm} \frac{x^i}{K_{\text{inter}}} \quad (6)$$

Because $[R_1] = [L_1] + [M_1]$, from eq 6, considering the molar fraction of L_1 and M_1 , eqs 7 and 8 are obtained

$$[L_1] = \frac{x}{lK_{\text{inter}}} \quad (7)$$

$$[M_1] = \frac{x}{mK_{\text{inter}}} \quad (8)$$

The assembly S is in general a polycyclic molecule whose degree of cyclicity is given by the minimum number of bonds that must be broken to transform S into an acyclic molecule. To join $pm L_1$ and $pl M_1$ molecules, $N - 1$ intermolecular bonds are required. Because the assembly of S requires in total B bonds, the number of intramolecular bonds, coinciding with the degree of cyclicity, is $B - N + 1$. The self-assembly equilibrium constant K_S is thus given by

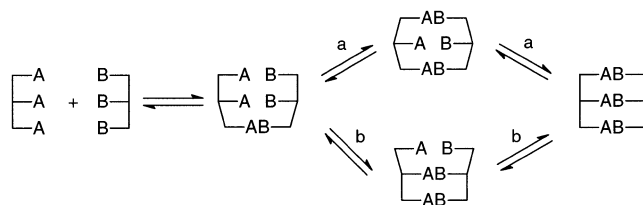
$$K_S = \sigma_r K_{\text{inter}}^{N-1} \prod_{i=1}^{B-N+1} K_{(\text{intra})i} \quad (9)$$

where σ_r is a statistical factor that accounts for the number of equivalent binding sites of the reactants. Considering that for L_1 and M_1 these are l and m , respectively, and taking into account the corresponding stoichiometric coefficients, it results that $\sigma_r = l^m m^p$.⁶ The equilibrium constants $K_{(\text{intra})i}$ refer to the intramolecular processes (cyclizations) required for the assembly of S . The cyclization tendency is usually expressed in terms of effective molarity (EM) defined as $K_{\text{intra}}/K_{\text{inter}}$.^{7,8} Introducing the expressions for σ_r and EM into eq 9, eq 10 is obtained

$$K_S = l^m m^p K_{\text{inter}}^{B-N+1} \prod_{i=1}^{B-N+1} EM_i \quad (10)$$

Note that the EM_i values are path dependent, but not their product. An example will suffice: consider the two pathways a and b in Scheme 2 for the self-assembly of a generic bicyclic product. Because the EM is strongly dependent on the ring size and on the possible strain of the ring being formed, it is apparent that the two pathways are characterized by different EM_i values; nevertheless the product of the EM_i of each of the two pathways obviously must be the same in that the two pathways lead to the same molecule.

SCHEME 2



To avoid considering a specific path, it is convenient to define the EM of the complex *S* as the geometrical mean of the EM_{*i*} values for an arbitrary path,⁹ so that eq 10 becomes

$$K_S = l^m m^p l^q K_{\text{inter}}^B \text{EM}^{B-N+1} \quad (11)$$

From the definition of K_S in terms of the concentrations of the species involved, and considering eqs 7, 8, and 11, one obtains

$$[S] = K_{\text{inter}}^{B-N} \text{EM}^{B-N+1} x^N \quad (12)$$

that substituted in eq 4 affords

$$[L_1]_0 + [M_1]_0 = \frac{1}{K_{\text{inter}}} \left(NK_{\text{inter}}^{B-N+1} \text{EM}^{B-N+1} x^N + \frac{N}{B} \frac{x}{(1-x)^2} \right) \quad (13)$$

Although eq 13 cannot be solved analytically, knowing the initial concentrations of the reactants, K_{inter} , and EM, it is possible to obtain a numerical solution for x by the Newton–Raphson method.¹⁰ In turn, knowledge of x allows the concentrations of the self-assembling complex and of the various families of oligomers to be found by eqs 12 and 6, respectively.

At this point, however, it is important to remark that the preceding equations are valid as long as all the species are present in solution. It is well-known that nonlinear polymerization, in contrast with linear polymerization, at a sufficiently advanced extent of reaction in the polymer fraction, produces indefinitely large polymer structures limited in size only by the macroscopic amount of material in the reaction mixture.⁵ This occurrence is signaled by a sharp transition of the solution from a liquid to an elastic gel (gel-point). The Flory–Stockmayer theory of gelation predicts that under the conditions of the present study (equal reactivity of like functional groups, absence of intramolecular reaction in the polymer, equal amounts of functional groups –A and –B) the gel-point occurs at a critical value of x given by⁵

$$x_c = \frac{1}{\sqrt{(l-1)(m-1)}} \quad (14)$$

Because the equations here obtained are valid under pre-gel conditions, eq 14 sets an upper limit for the value of x . In fact, this is not a limitation of the model because beyond the gel-point, self-assembly could not effectively compete with gel formation.

The percent weighted fraction (the yield) of the assembly, given by $100M[S]/([L_1]_0 + [M_1]_0)$, is expressed, taking into account eqs 12 and 13, by

$$\text{S\%} = 100 \left(1 + \frac{1}{BK_{\text{inter}}^{B-N+1} \text{EM}^{B-N+1} (1-x)^2 x^{N-1}} \right) \quad (15)$$

Equations 13 and 15 can be used to build distribution curves of S% vs $K_{\text{inter}}([L_1]_0 + [M_1]_0)$, that depend only on the product

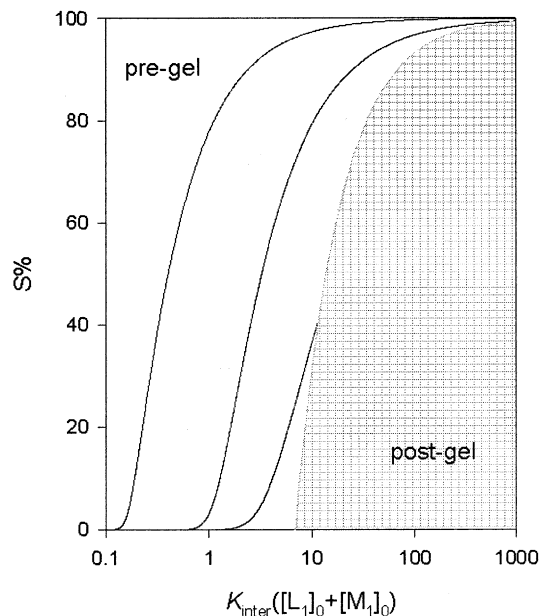


Figure 2. Plots of the yield of the self-assembling cube sketched in Figure 1 against $K_{\text{inter}}([L_1]_0 + [M_1]_0)$. The three curves from the bottom up, refer to $K_{\text{inter}}\text{EM} = 3, 10,$ and $500,$ respectively. The curves are defined in the pre-gel area ($0 < x < 0.707$).

$K_{\text{inter}}\text{EM}$. For example, in Figure 2 are shown plots regarding the self-assembly of the cube in Figure 1 ($N = 20$ and $B = 24$) for values of $K_{\text{inter}}\text{EM} = 3, 10,$ and $500,$ respectively.¹¹ The plots are defined up to the gel-point occurring, according to eq 14, at $x_c = 0.707$.

From eq 14 it appears that the driving force for the formation of *S* is given by the term $(K_{\text{inter}}\text{EM})^{B-N+1}$ that is strongly dependent on the degree of cyclicity of *S*. In the example shown in Figure 2 the degree of cyclicity is rather high, namely 5, and this makes the yield of *S* highly sensitive to the variation of $K_{\text{inter}}\text{EM}$. It is apparent that an increase of $K_{\text{inter}}\text{EM}$ is beneficial not only to the yield of the assembly but also to the concentration range over which self-assembly takes place. The value of K_{inter} by itself affects the value of the monomers concentration at which a given yield of *S* is obtained; higher values of K_{inter} make the phenomenon observable at higher dilution.

Examination of eq 15 shows that S% tends to zero at very high dilution where only the monomers are present in solution ($x \rightarrow 0$); on increasing x , S% increases up to a maximum and then decreases because of the competition with polymerization. The maximum of eq 15 occurs at a value of x given by

$$x_{\text{Smax}} = \frac{N-1}{N+1} \quad (16)$$

Now we are in the position to know the minimum value of the driving force, $K_{\text{inter}}\text{EM}$, required for virtually complete self-assembly ($\text{S\%} \geq 99.01$). According to eq 15, this condition is satisfied when

$$K_{\text{inter}}\text{EM} \geq \left[\frac{100}{B(1-x)^2 x^{N-1}} \right]^{1/(B-N+1)} \quad (17)$$

By comparing eq 14 and eq 16, two cases can occur: (i) $x_{\text{Smax}} \leq x_c$; (ii) $x_{\text{Smax}} > x_c$.

Case i. In this case the maximum of S% occurs under pre-gel conditions; therefore the minimum value of the driving force

for complete self-assembly is obtained by substitution of eq 16 into eq 17, to give

$$K_{\text{inter}}\text{EM} \geq \left(100 \frac{aN^2}{B}\right)^{1/(B-N+1)} \quad (18)$$

where

$$a = \frac{(N+1)^{N+1}}{4N^2(N-1)^{N-1}}$$

Because a is confined in the very narrow range $1.69 \leq a < e^2/4$ ($= 1.85$) for $2 \leq N < \infty$, the condition in eq 18 can be translated in the useful form

$$K_{\text{inter}}\text{EM} \geq \left(185 \frac{N^2}{B}\right)^{1/(B-N+1)} \quad (19)$$

Equation 19 holds also for the case of two-dimensional self-assembly accompanied by linear polymerization, in which gelation does not occur. Indeed, in this case $l = m = 2$ and $N = B$, and thus eq 19 reduces to eq 20, showing that the needed driving force directly depends on the number of bonds that hold together the monomer units in the cyclic assembly.

$$K_{\text{inter}}\text{EM} \geq 185B \quad (20)$$

This result had been previously obtained by a model limited to the treatment of self-assembly macrocyclizations,^{3a} which actually turns out to be a specific case of the present general model.

Case ii. In this case the maximum value of $S\%$ is attained in correspondence of the gel-point; therefore the minimum value of the driving force for complete self-assembly is obtained by substitution of eq 14 into eq 17, to give

$$K_{\text{inter}}\text{EM} \geq \left[\frac{100}{B} \frac{b^{N+1}}{(b-1)^2}\right]^{1/(B-N+1)} \quad (21)$$

where $b = 1/x_c$.

Both eqs 19 and 21 are important relations that allow a prediction of the feasibility of self-assembly on the basis of simple structural parameters such as the number of molecules in the assembly, the number of bonds joining them, and the number of interaction sites in the monomers. For example, consider the cube in Figure 1, for which $x_c = 0.707$ and $x_{S\text{max}} = 0.905$. Because $x_{S\text{max}} > x_c$, the condition in eq 21 holds, and a value of $K_{\text{inter}}\text{EM} > 8.1$ is calculated. A close inspection of Figure 2 indeed confirms that only the curves relative to $K_{\text{inter}}\text{EM} = 10$ and 500 show complete self-assembly.

Both eqs 19 and 21 point out that *the required driving force for self-assembly decreases on decreasing the number of molecules of the assembly and on increasing the number of bonds joining them*. Although analogous conclusions have been intuited by other authors,^{2b,12} it is the first time that quantitative relationships are obtained on the basis of a rigorous model. It should be clear that the term $(K_{\text{inter}}\text{EM})^{B-N+1}$ is simply the equilibrium constant between the family of acyclic oligomers of molecular formula $L_{pm}M_{pl}$ and the assembly S . In free energy terms, $(K_{\text{inter}}\text{EM})^{B-N+1} = \exp(-\Delta G^\circ/RT)$ where $-\Delta G^\circ$ is the difference between the standard free energy in solution of S and $L_{pm}M_{pl}$. Thus eqs 19 and 21 simply express the conditions about the stability of S so that it can win the competition with all the acyclic oligomers.

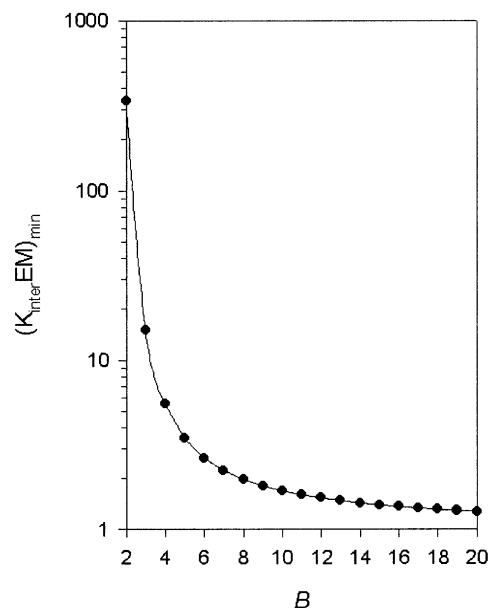


Figure 3. Plot of $(K_{\text{inter}}\text{EM})_{\text{min}}$ for $N = 2$ vs B . The plot can be referred to the pairing of two strands having B interaction sites. The first three points were calculated by eq 18 and the remainder by eq 21.

Both eqs 19 and 21 put also in quantitative form the assumption, made at the beginning, that the monomers have a rigid structure predisposed to form a given assembly in preference to other hypothetical cyclic or polycyclic species. This assumption holds when the EM of the assembly satisfies either eq 19 or eq 21 whereas the EMs of all other hypothetical assemblies are significantly below the limits established by these equations. In general, it is difficult to compare the stabilities of assemblies made from the same building blocks but differing in the molecular formula. In this respect eqs 19 and 21 can provide a useful insight in that only the complex satisfying the appropriate equation can exclusively form. There could be the case in which more than one assembly satisfies either eq 19 or eq 21. In general, lower order assemblies are stable to lower monomer concentrations, so that there could be distinct concentration ranges in which each assembly exclusively forms, as well as a concentration range in which the assemblies are both present in significant amounts. Examples of the latter case have been reported in the literature, especially as far as “triangle-square” equilibria are concerned.^{2i,13}

The self-assembly process requiring the lowest driving force is that involving a large number of bonds with respect to the number of molecules involved, a typical example is the pairing of two complementary strands having a large number of interaction sites along the chain. In Figure 3 is plotted the minimum value of $K_{\text{inter}}\text{EM}$ for $N = 2$ vs B . It appears that on increasing B , there is initially a sharp drop of the required driving force for complete self-assembly followed by a leveling off. Independently from the number of molecules constituting the assembly, on increasing B , eq 21 tends to the limit $K_{\text{inter}}\text{EM} > 1$. This is an interesting result because it indicates that no matter how large the number of bonds holding together the pieces of an assembly, there is a threshold value of the driving force that if not exceeded makes self-assembly impossible.

Another interesting aspect regards the sensitivity with which the yield of an assembly changes on changing its degree of cyclicity. Indeed, because a high degree of cyclicity of S involves its yield being highly sensitive to the variation of $K_{\text{inter}}\text{EM}$, for an assembly with a great number of interaction sites the limit established by eq 21 becomes the boundary

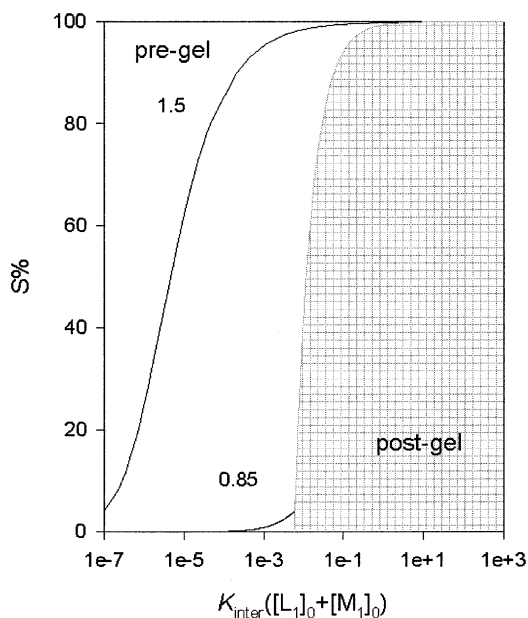


Figure 4. Plots of the yield of a duplex formed by pairing of two strands with 20 binding sites against $K_{\text{inter}}([L]_0 + [M]_0)$ for $K_{\text{interEM}} = 0.85$ and 1.5.

between an all-or-nothing process. To illustrate this point, in Figure 4 are reported two distribution curves referring to the self-assembly of two strands with 20 interaction sites per monomer to form a duplex. From eq 21 it results that complete self-assembly requires $K_{\text{interEM}} > 1.3$. The two curves refer to $K_{\text{interEM}} = 0.85$ and 1.5. It is evident that a modest change of the value of K_{interEM} about the boundary can cause a dramatic effect on the formation of the assembly. Depending on the value of the monomers concentration, the assembly could undergo either sharp disassembly (*denaturation*) or conversion into gel. Variations of K_{interEM} can be caused by variations of temperature, ionic strength, solvent composition, etc. that act on the value of K_{inter} , because the EM is known to be rather insensitive to medium variations as well as, in the absence of strain, to temperature.^{3b,8}

Conclusion

A general model for self-assembly in solution has been presented in which the process occurs in competition with nonlinear random polymerization. The model is based on two assumptions: (i) the monomers have a rigid structure predisposed to form a given assembly in preference to other hypothetical cyclic or polycyclic species; (ii) all the intermolecular reactions between the binding sites occur with the same probability irrespective of the oligomer to which they are attached and the position they occupy within a given oligomer. The fundamental quantities on which the treatment is based are the effective molarity (EM) of the assembly and the equilibrium constant for the intermolecular model reaction between monofunctional reactants (K_{inter}). The condition for complete self-assembly is expressed either by eq 19 or eq 21, both stating that the product K_{interEM} must be no lower than an easily computable limit that depends on the number of molecules constituting the assembly (N), the number of bonds (B) holding them together, and the number of interaction sites in the monomers. This limit decreases on decreasing N and on increasing B , and the most obvious way to realize this condition is by increasing the degree of cyclicality of the assembly. It is evident that to exploit the power of the above equations, it is

crucial to have methods that allow an estimate of both K_{inter} and EM. In fact, K_{inter} is easily available either by studying the equilibrium between monofunctional reactants or on the basis of literature data, whereas an estimate of the EM is more difficult. The EM has both an enthalpic and an entropic component. The factors that possibly affect the enthalpic component are the strain energy of the ring structure (that decreases the EM), intra-annular stabilizing interactions, and template effects (both increasing the EM). All of these effects can be absent, in this case the EM solely depends on entropy. Essentially, the following factors affect the entropic component: (i) The intrinsic entropic advantage of an intramolecular reaction with respect to the intermolecular counterpart due to the significant losses of translational plus rotational entropy that occur in the latter. According to Page and Jencks, entropy changes from -35 to -50 eu ($1 \text{ eu} = 4.184 \text{ J K}^{-1} \text{ mol}^{-1}$) are to be expected for many bimolecular reactions in solution at 25°C . This corresponds to a maximum entropic advantage of about 10^8 mol L^{-1} for a comparable intramolecular reaction.^{8a,b} (ii) The loss of entropy due to the restriction of torsional motion in the cyclic structure. (iii) The number of symmetry, σ , of the formed ring, corresponding to the number of equivalent bonds that can be broken in the reverse reaction, which decreases the entropy by $R \ln \sigma$.⁶ In a previous paper we have suggested a method, based on molecular modeling techniques, to obtain a lower limit for the EM of a fairly rigid two-dimensional assembly.^{3b} Other empirical methods have been suggested in the case of strainless rings.^{8d,e} Considering that in the case of polycyclic assemblies, the EM is the geometrical mean of $B - N + 1$ effective molarities, these methods could be generalized to obtain a rough estimate of the EM value of a three-dimensional complex. In many cases this is sufficient to establish if the appropriate equation, eq 19 or eq 21, is satisfied and, therefore, if self-assembly can be complete.

It is useful to recall that maximization of the EM requires a careful design of the monomeric building blocks. Ideal building blocks are rigid (no built-in rotors) and have a structural predisposition to assemble in a unique supramolecular structure without any strain. The presence of templates and/or intraannular stabilizing interactions can greatly enhance the EM. Moreover, in the course of the formation of a polycyclic structure, the ring closure of a successive ring can be favored by the formation of the preceding rings. This effect, involving positive cooperativity, has been evidenced in the self-assembly of DNA,^{2d} and of artificial supramolecular structures.¹⁴ In the example of two strands that self-assemble to form a duplex, positive cooperativity would manifest itself by an increase of the EM on increasing the number of binding sites along the strands.

The ideal intermolecular reaction, also expressed in the structure of the building blocks, involves the formation of nonrotatable bonds. This can be realized by reactions involving either multipoint binding (for example, multiple hydrogen bonding), or steric hindrance to rotation of the newly formed bonds, or the formation of multiple bonds (for example, reactions on transition metals involving back-donation, formation of imines, etc.). The ideal intermolecular reaction, of course, is also highly esoergonic to maximize the value of K_{inter} . The degree of cyclicality of the supramolecular structure, although it does not affect the EM in the absence of cooperativity, should be as high as possible so as to minimize the required driving force.

One may wonder if self-assembly can be driven solely by entropy. The driving force for self-assembly is given by the product K_{interEM} , and although the EM can be purely entropic,

K_{inter} referring to the association of the groups $-A$ and $-B$ can hardly be entropically favored unless a large number of solvent molecules are released upon association. However, if the following conditions are satisfied, (i) the EM is purely entropic, (ii) $K_{\text{inter}} \leq 1 \text{ M}^{-1}$, and (iii) the product $K_{\text{inter}}EM$ satisfies either eq 19 and 21, then self-assembly is driven by entropy only.

In conclusion, knowledge of EM and K_{inter} allows the evaluation of the distribution curve of the assembly, and conversely, from the distribution curve, these fundamental physicochemical parameters can be obtained.^{3b} It is desirable that besides structural data aimed at characterizing self-assembling architectures, more data about their robustness in solution, as quantitatively expressed by EM and K_{inter} , will be available in the future.

References and Notes

- (1) Part 4 of the series Physical Basis of Self-Assembly. Part 3: Ercolani, G. *J. Chem. Soc., Chem. Commun.* **2001**, 1416–1417.
- (2) For recent reviews see: (a) Lehn, J.-M. *Supramolecular Chemistry: Concepts and Perspectives*; VCH Publishers: Weinheim, 1995. (b) Whitesides, G. M.; Simanek, E. E.; Mathias, J. P.; Seto, C. T.; Chin, D. N.; Mammen, M.; Gordon, D. M. *Acc. Chem. Res.* **1995**, *28*, 37–44. (c) Lawrence, D. S.; Jiang, T.; Levett, M. *Chem. Rev.* **1995**, *95*, 2229–2260. (d) Philp, D.; Stoddart, J. F. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1154–1196. (e) *Comprehensive Supramolecular Chemistry*; Atwood, J. L., Davies, J. E. D., MacNicol, D. D., Vögtle, F., Sauvage, J.-P., Hosseini, M. W., Eds.; Pergamon: Oxford, 1996; Vol. 9. (f) Zeng, F. W.; Zimmerman, S. C. *Chem. Rev.* **1997**, *97*, 1681–1712. (g) Linton, B.; Hamilton, A. D. *Chem. Rev.* **1997**, *97*, 1669–1680. (h) Conn, M. M.; Rebek, J. *Chem. Rev.* **1997**, *97*, 1647–1668. (i) Fujita, M. *Chem. Soc. Rev.* **1998**, *27*, 417–425. (j) Caulder, D. L.; Raymond, K. N. *J. Chem. Soc., Dalton Trans.* **1999**, 1185–1200. (k) Leininger, S.; Olenyuk, B.; Stang, P. J. *Chem. Rev.* **2000**, *100*, 853–908. (l) Lindoy, L. F.; Atkinson, I. In *Self-Assembly in Supramolecular Systems (Monographs in Supramolecular Chemistry)*; Stoddart, J. F., Ed.; The Royal Society of Chemistry: Cambridge, U.K., 2000. (m) *Molecular Self-Assembly*; Fujita, M., Ed.; Structure and Bonding, Vol. 96; Springer: Heidelberg, 2000. (n) Swiegers, G. F.; Malefetse, T. J. *Chem. Eur. J.* **2001**, *7*, 3636–3643. (o) Albrecht, M. *Chem. Rev.* **2001**, *101*, 3457–3497. (p) Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. *Angew. Chem., Int. Ed. Engl.* **2002**, *41*, 898–952. (q) Seidel, S. R.; Stang, P. J. *Acc. Chem. Res.* **2002**, *35*, 972–983.
- (3) (a) Ercolani, G. *J. Phys. Chem. B* **1998**, *102*, 5699–5703. (b) Ercolani, G.; Ioele, M.; Monti, D. *New J. Chem.* **2001**, *25*, 783–789. (c) Bielejewska, A. G.; Marjo, C. E.; Prins, L. J.; Timmerman, P.; de Jong, F.; Reinhoudt, D. N. *J. Am. Chem. Soc.* **2001**, *123*, 7518–7533.
- (4) The term predisposition has been used in this context to indicate a strong conformational or structural preference expressed by the building block once incorporated into a larger structure, giving rise to a thermodynamic preference for a particular product. Rowan, S. J.; Hamilton, D. G.; Brady, P. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1997**, *119*, 2578–2579.
- (5) For leading references, see: (a) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1953; Chapter 9. (b) Stepto, R. F. T. In *Developments in Polymerisation/3*; Harward, R. N., Ed.; Applied Science Publishers: London, 1982; Chapter 3. (c) *Polymer Networks: Principles of their Formation, Structure and Properties*; Stepto, R. F. T., Ed.; Blackie Academic and Professional: London, 1998. (d) Stepto, R. F. T.; Taylor, D. J. R. In *Cyclic Polymers*, 2nd ed.; Semlyen, J. A., Ed.; Kluwer: Dordrecht, The Netherlands, 2000; Chapter 15.
- (6) The statistical factor of a molecule is equivalent to its external symmetry number, σ , defined as the number of indistinguishable but nonidentical positions into which the molecule can be turned by rigid rotation. It affects the entropy of the molecule by a factor $-R \ln \sigma$ (Eliel, E. L.; Wilen, S. H. *Stereochemistry of Organic Compounds*; Wiley: New York, 1994; pp 96 and 601). Thus for a generic equilibrium $aA + bB \rightleftharpoons cC$, if σ_A , σ_B , and σ_C are the symmetry numbers of A, B, and C, respectively, the effect of their symmetry on the equilibrium constant is given by the factor $\sigma_A^a \sigma_B^b / \sigma_C^c$. The effect of the number of symmetry of S on the self-assembly equilibrium, corresponding to the number of equivalent bonds that can be broken in the reverse reaction, is not explicitly considered, because this effect is incorporated in the values of $K_{\text{intra}i}$.
- (7) It has been shown (ref a) that the EM is both conceptually and operationally identical to the molar cyclization equilibrium constant K introduced by Jacobson and Stockmayer in their classical paper on linear ring-chain equilibria (ref b). (a) Ercolani, G.; Mandolini, L.; Mencarelli, P.; Roelens, S. *J. Am. Chem. Soc.* **1993**, *115*, 3901–3908. (b) Jacobson, H.; Stockmayer, W. H. *J. Chem. Phys.* **1950**, *18*, 1600–1606.
- (8) For reviews on the concept of EM and its applications to ring closure reactions, see: (a) Page, M. I.; Jencks, W. P. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1678–1683. (b) Page, M. I. *Chem. Soc. Rev.* **1973**, *2*, 295–323. (c) Kirby, A. J. *Adv. Phys. Org. Chem.* **1980**, *17*, 183–278. (d) Mandolini, L. *Adv. Phys. Org. Chem.* **1986**, *22*, 1–111. (e) Galli, C.; Mandolini, L. *Eur. J. Org. Chem.* **2000**, 3117–3125.
- (9) An analogous definition of EM has been already given by other authors in the case of the self-assembly of C_3 symmetry molecular cages: Felluga, F.; Tecilla, P.; Hillier, L.; Hunter, C. A.; Licini, G.; Scrimin, P. *J. Chem. Soc., Chem. Commun.* **2000**, 1087–1088.
- (10) Margenau, H.; Murphy, G. M. *The Mathematics of Physics and Chemistry*, 2nd ed.; Van Nostrand Company: Princeton, NJ, 1962; p 492.
- (11) The plots in Figure 2 can be built in two ways: (i) For each $K_{\text{inter}} - ([L_1]_0 + [M_1]_0)$ value, the equation obtained from eq 13 multiplying both sides by K_{inter} , is solved for x by the Newton–Raphson method, and then, knowing x , S% can be calculated by eq 15. Points referring to x values greater than 0.707 are rejected. (ii) A list of x values in the range $0 \leq x \leq 0.707$ is compiled, and then the couple of $K_{\text{inter}}([L_1]_0 + [M_1]_0)$ and S% values corresponding to each x is calculated by the equation obtained from eq 13 by multiplying both sides by K_{inter} and from eq 15, respectively. The second method is simpler and much faster than the first one.
- (12) (a) Hunter, C. A. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1079–1081. (b) Chi, X.; Guerin, A. J.; Haycock, R. A.; Hunter, C. A.; Sarson, L. D. *J. Chem. Soc., Chem. Commun.* **1995**, 2563–2565.
- (13) (a) Fujita, M.; Sasaki, O.; Mitsuhashi, T.; Fujita, T.; Yazaki, J.; Yamaguchi, K.; Ogura, K. *J. Chem. Soc., Chem. Commun.* **1996**, 1535. (b) Fujita, M.; Ogura, K. *Bull. Chem. Soc. Jpn.* **1996**, *69*, 1471–1482. (c) Fujita, M.; Ogura, K. *Coord. Chem. Rev.* **1996**, *148*, 249–264. (d) Lee, S. B.; Hwang, S.; Chung, D. S.; Yun, H.; Hong, J.-I. *Tetrahedron Lett.* **1998**, *39*, 873–876. (e) McQuillan, F. S.; Berridge, T. E.; Chen, H.; Hamor, T. A.; Jones, C. J. *Inorg. Chem.* **1998**, *37*, 4959–4970. (f) Sun, S.-S.; Lees, A. J. *Inorg. Chem.* **1999**, *38*, 4181–4182. (g) Sun, S.-S.; Lees, A. J. *J. Am. Chem. Soc.* **2000**, *122*, 8956–8967. (h) Schnebeck, R.-D.; Freisinger, E.; Lippert, B. *Eur. J. Inorg. Chem.* **2000**, 1193–1200. (i) Sautter, A.; Schmid, D. G.; Jung, G.; Würthner, F. *J. Am. Chem. Soc.* **2001**, *123*, 5424–5430.
- (14) (a) Pfeil, A.; Lehn, J.-M. *J. Chem. Soc., Chem. Commun.* **1992**, 838–840. (b) Garrett, T. M.; Koert, U.; Lehn, J.-M. *J. Phys. Org. Chem.* **1992**, *5*, 529–532. (c) Taylor, P. N.; Anderson, H. L. *J. Am. Chem. Soc.* **1999**, *121*, 11538–11545.