Symmetry Numbers and Statistical Factors in Self-Assembly and Multivalency

Gianfranco Ercolani,*,† Claude Piguet,‡ Michal Borkovec,‡ and Josef Hamacek‡

Dipartimento di Scienze e Tecnologie Chimiche, Università di Roma Tor Vergata, Via della Ricerca Scientifica, 00133 Roma, Italy, and Department of Inorganic, Analytical and Applied Chemistry, University of Geneva, 30 quai E. Ansermet, 1211, Geneva 4, Switzerland

Received: May 25, 2007; In Final Form: August 3, 2007

Evaluation of statistical factors in self-assembly processes is not a firmly settled question. As a contribution to solve this problem, a critical re-examination of the symmetry number method and generalization of the direct count method are presented. The two approaches, producing the same results, mutually reinforce their role with respect to other discordant methods whose results cannot be independently checked. The direct count method moreover serves as a rationale for the apparently odd results the symmetry number method sometimes provides. The two methods thus turn out to be complementary to each other. Discussion of some exemplary cases points to the importance and subtlety of the role played by the geometrical features of assemblies involving intramolecular bonds.

Introduction

Multivalent noncovalent interactions are essential ingredients in the mediation of biological processes, as well as in the self-assembly of functional nanosystems for material applications. A fundamental understanding of the organizing principles of multivalency[1] and self-assembly[2] is therefore of paramount importance, not only for the construction of complex (super)-structures and devices, but also to provide insight into how biological processes work.

In recent years, a number of methods have been proposed to factorize the stability constant of an assembly into inter- and intramolecular free-energy contributions and an entropic (statistical) contribution accounting for the degeneracy of bound states.[3−6] Although all the proposed methods share the latter factor, they differ in the suggested procedure for its evaluation, often providing equivocal and inconsistent results that may engender controversy.[4b] In fact, an accurate and consistent evaluation of statistical factors in self-assembly processes is of great importance to predict the expected stability constant in the absence of cooperative effects and, therefore, to spotlight the emergence of either positive or negative cooperativity as a marked deviation from statistical behavior.

The proposed approaches to evaluate statistical factors are based either on the evaluation of symmetry numbers of reactants and products,[4] or on different flavors of combinatorial procedures.[3,5,6] Here we argue that, in the case of conflict, the symmetry number method, pioneered by Benson,[7,8] provides the correct results. In particular, to provide a rationale for the odd results sometimes provided by the symmetry number method, here we wish to illustrate the generalization of a previously published method,[9] aimed at the direct count of the number of distinct microspecies of an assembly. Thus the two methods mutually reinforce their role with respect to other discordant methods whose results cannot be independently checked.

For the sake of clarity, the article is broken down into two main sections: Theory and Discussion. In the first part, the symmetry number method and the direct count method are critically re-examined under separate headings. In the second part, application of the theory to both hypothetical and real cases is discussed, with particular emphasis to the comparison of assemblies without and with intramolecular bonds.

Theory

The Symmetry Number Method. According to Benson,[7,8a] the observed equilibrium constant, K, of a generic equilibrium (eq 1) can be regarded as being given by the product of an intrinsic or "chemical" constant K_{chem} and a statistical factor, K_σ, given by the ratio of symmetry numbers for reactant and product species in equilibrium (eq 2).

$$a\mathrm{A} + b\mathrm{B} \overset{K = K_\sigma K_{chem}}{\rightleftharpoons} c\mathrm{C} + d\mathrm{D} \tag{1}$$

$$K_\sigma = \frac{\sigma_{reactants}}{\sigma_{products}} = \frac{\sigma_A{}^a \sigma_B{}^b}{\sigma_C{}^c \sigma_D{}^d} \tag{2}$$

Equation 2 is derived from the fact that the symmetry number of a molecule, σ, affects its rotational entropy by a factor $-R \ln \sigma$. The factor σ is the product of the external (σ_{ext}) and internal (σ_{int}) symmetry numbers. The external symmetry number is defined as the number of different but indistinguishable atomic arrangements that can be obtained by rotating a given molecule as a whole. In practice, it is found by multiplying the order of the independent simple rotational axes of the point group to which the molecule belongs (in this respect, axes of infinite order are not considered because they do not generate different atomic arrangements). External symmetry numbers for the various point groups are shown in Table 1.

The internal symmetry number is similarly defined as the number of different but indistinguishable atomic arrangements that can be obtained by internal rotations around single bonds. There is a controversy here: according to some authors,[8b,10] the use of the internal symmetry number is correct only for

* Corresponding author. E-mail: ercolani@uniroma2.it.
† Università di Roma Tor Vergata.
‡ University of Geneva.

10.1021/jp0740705 CCC: $37.00 © 2007 American Chemical Society
Published on Web 10/05/2007

TABLE 1: External Symmetry Numbers for Various Point Groups

point group	σ_{ext}
C_1, C_i, C_s, $C_{\infty v}$, R_3	1
$D_{\infty h}$	2
C_n, C_{nv}, C_{nh}	n
D_n, D_{nd}, D_{nh}	$2n$
S_n (n even)	$n/2$
T_d	12
O_h	24
I_h	60

free or practically free internal rotations (energy barrier, $V_0 \ll RT \approx 0.6$ kcal mol^{-1} at room temperature), whereas others also use the internal symmetry number for barriers significantly higher than RT without any comment about the height of the barrier.[8a] For example, according to Benson, ethane has $\sigma = 18$, which is the product of $\sigma_{ext} = 6$ because of D_{3d} symmetry and $\sigma_{int} = 3$, in spite of $V_0 \approx 3$ kcal mol^{-1}. In our opinion, provided that the internal rotation is fast with respect to the time scale in which the equilibrium in eq 1 is attained and measured, the complete potential energy curve encompassing all the degenerate minima should be taken into account for the quantum mechanical evaluation of the partition function for internal rotation. When this is done, the partition function for hindered internal rotation, independently of the height of the barrier, is inversely proportional to σ_{int}[11] as the partition function for free internal rotation, thus justifying the approach of Benson.

A problem analogous to internal rotation is encountered in the case of fluxional molecules such as NH_3, cyclobutane, and so forth, in which two or more chemically equivalent configurations are separated by an energy barrier. By analogy with σ_{int}, we propose the introduction of a fluxional symmetry number, σ_{flux}, defined as the number of different but indistinguishable atomic arrangements that can be obtained by inversion, pseudorotation, or other intramolecular processes. Also, in this case, the quantum mechanically evaluated partition function accounting for all the degenerate minima separated by an identical energy barrier depends on the inverse of σ_{flux}, independently of the height of the barrier with respect to RT. The problem was discussed by Pitzer in the case of ammonia inversion:

> Although the question as to how rapid reversing is needed to change the symmetry number [from 3 to 6] would seem troublesome, it turns out to be otherwise. Thus one finds that for any height of central potential peak one may consider the possibility of reversal taking the larger symmetry number and the vibration energy levels [relative to the two minima].[12]

A different type of correction accompanies chemically distinguishable isomeric species at equilibrium with each other that are treated collectively as a single compound.[13] They can be constitutional isomers, stereoisomers, rotational conformers, and so forth. Let us suppose that the various isomers have the same free energy content apart from the entropy of symmetry, thus a correction accounting for the entropy of mixing of these species is needed. A way to handle this problem is to consider an apparent symmetry number for the whole isomeric population. Indeed, the overall free energy of an equilibrium mixture of isomers is given by eq 3, where i runs over all isomers, each characterized by its own free energy G_i^0.[14]

$$G^0 = -RT \ln \sum_i e^{-G_i^0/RT} \tag{3}$$

The advantage of eq 3 with respect to the usual formula for the entropy of mixing[8] is that it implicitly considers such a source of entropy without the need to calculate the molar fractions of the components of the mixture. Since the relative equilibrium concentrations of the isomers are dictated by their symmetry only, the excess free energy of each isomer is $G_i^0 = RT \ln \sigma_i$. By introducing this expression in eq 3, the excess free energy of the isomeric population is

$$G^0 = -RT \ln \sum_i \frac{1}{\sigma_i} \tag{4}$$

By defining an apparent symmetry number for the whole isomeric mixture such that $G^0 = RT \ln \sigma_{app}$, and equating this expression to eq 4, eq 5 is obtained:

$$\sigma_{app} = \frac{1}{\sum_i \frac{1}{\sigma_i}} \tag{5}$$

If all the n components of the mixture have the same symmetry number, say σ, eq 5 reduces to eq 6, in which the effect due to the entropy of mixing ($1/n$) is separable from σ.

$$\sigma_{app} = \frac{\sigma}{n} \tag{6}$$

An important and frequent case requiring the application of eq 6 is that of a chiral molecule present at equilibrium as a racemic mixture. Two enantiomers have exactly the same free energy and the same symmetry number. Thus if a molecule is chiral and is present at equilibrium as a racemic mixture, its symmetry number must be divided by 2 to account for the entropy of mixing of the two enantiomers.

In light of the discussion above, the evaluation of the symmetry number for rigid molecules, possibly including the presence of internal rotors, fluxional processes, and the formation of isoenergetic isomers, is quite straightforward. Some complication arises in the case of conformationally mobile molecules giving rise to conformers of different energy. Although, in principle, an apparent symmetry number for such an isomeric mixture could be obtained, the procedure would require a detailed knowledge of the free energy of all the conformers, which in most of the cases is not available. In order to devise a simple and useful approximation to handle this problem in a consistent way, it is important to realize that the factorization of an equilibrium constant, K, in the product $K_\sigma K_{chem}$ involves a certain degree of arbitrariness that depends on the level of knowledge about the system at hand. For example, consider a hypothetical equilibrium in which a molecule of H_2O_2 is involved as a reactant. H_2O_2 has an external twofold axis; moreover, the internal rotation presents two distinct minima, mirror images of each other, in which the O−H bonds are nearly at right angles; thus, applying eq 6, $\sigma_{app} = 1$. An investigator, ignorant of the existence of the two minima, might assume that the internal rotation is free and conclude that $\sigma = 2$. In this case, the values of K_σ and K_{chem} are twice and half, respectively, of those of the more informed investigator; however, as long as this result is used systematically for all the reactions involving H_2O_2, the conclusions obtained by the unknowing investigator would be fully consistent. We wish, now, to discuss the archetypal cases of linear and cyclic alkanes in some detail. Benson, after having discussed the symmetry number of ethane, added "On the same basis all the normal alkanes have $\sigma = 18$, due to an external twofold axis, and $\sigma_{int} = 3^2$ for the two

terminal CH_3 groups".[8a] Apart from the case of propane, for longer alkanes, this analysis is rather superficial. In fact, it would be correct only if the chains were exclusively present in the extended conformation. Although this conformation is the global minimum, its stability is not so large that it excludes the significant presence of other distinct conformations as described by the *rotational isomeric state model*.[15] As pointed out above, however, evaluation of the apparent symmetry number for such an isomeric mixture would be unduly complicated in most of the cases. Accordingly, we purposely adopt the crude approximation that considers all the rotations about the skeletal bonds of the chain as free, meaning that all torsion angles have equal probability of occurrence.[16] This view implies that the various conformations are chemically indistinguishable from each other, a notion that is in keeping with the common experience of chemists. For butane, it is immediately obvious that, within the framework of the free rotor approximation, $\sigma = 18$ because butane has an external twofold axis independently of the value of the torsional angle between the central carbons. It is not so obvious that, also in the case of longer alkanes, $\sigma = 18$, although it can be demonstrated as follows. Suppose that like atoms in a molecule could be distinguished by labels. Then, if the molecule contains n_1 identical atoms of kind 1, ..., n_i of kind i, from one equilibrium configuration of the labeled molecule others could be made by permutation of like atoms, with the total number of configurations generated in this way being $\prod_i n_i!$. Some of these configurations might be related to others by rotation of the molecule as a whole, by internal rotations, or by fluxional processes. If the number of distinguishable *label*-isomers (isomers created by the labeling procedure making like atoms distinguishable) is Γ, then the molecule's symmetry number, σ, will be given by eq 7:

$$\sigma = \frac{\prod_i n_i!}{\Gamma} \tag{7}$$

For example, since methane has 24 different permutations and just 2 distinct *label*-enantiomers, $\sigma = 12$. In the case of a linear alkane with n carbons, there are $2n + 2$ hydrogens, thus the number of permutations is $n!(2n + 2)!$. The number of *label*-isomers is calculated as follows. The number of distinct *label*-chains of n carbons is $n!/2$ because a sequence of labeled carbons is identical if read from left to right or vice versa; the number of ways one can associate $2n + 2$ numbered hydrogens to a numbered chain of n carbons by assigning 3 hydrogens to the first carbon, 3 hydrogens to the last, and 2 hydrogens to each of the remaining $n - 2$ carbons is $(2n + 2)!/[(3!)^2\ (2!)^{n-2}]$, and, finally, since each carbon of the chain is a *label*-chiral center, the number of *label*-stereoisomers is 2^n. The number of *label*-isomers will be given by the product of the three terms above, that is, $n!(2n + 2)!/18$. Thus, by eq 7, $\sigma = 18$ for all the linear alkanes. As to cycloalkanes, apart from cyclopropane, which is really planar, fast conformational motion makes all the $2n$ hydrogens equivalent as if the ring were planar, so a symmetry number of $2n$ could be guessed for all cycloalkanes.[17] Now we make the primitive assumption that the various accessible conformations are indistinguishable from each other and use eq 7 to evaluate the symmetry number of cycloalkanes. The general formula of a cycloalkane is C_nH_{2n}, thus the number of permutations is $n!2n!$. The number of *label*-isomers is calculated as follows. Considering a ring of n numbered carbons, ring opening by fission of one of the n C−C bonds generates one of the permutations of the linear chain; thus, from one numbered ring of n carbons, n linear permutations can be generated depending on the bond undergoing fission. Since the distinct permutations of a linear chain are $n!/2$, the distinct permutations of a ring chain will be $(n - 1)!/2$. The number of ways one can associate $2n$ numbered hydrogens in groups of 2 to each numbered carbon of the ring is $2n!/(2!)^n$. Finally considering that each carbon is a *label*-chiral center, 2^n *label*-stereoisomers are possible. The number of *label*-isomers, given by the product of the three terms, is $(n - 1)!2n!/2$, which, when introduced in eq 7, gives $\sigma = 2n$ for all cycloalkanes. These examples suggest that, in the case of flexible molecules, the approximation of indistinguishability between the various conformations leads to a consistent symmetry number that could be more easily obtained by considering the symmetry number of the most symmetric among the indistinguishable conformations, even though, as in the case of planar cycloalkanes (D_{nh}), this form does not even correspond to a minimum.

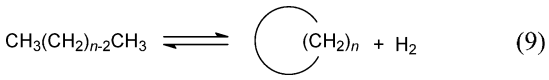
The approximation of indistinguishability for energetically different conformations of flexible chains makes the symmetry number method virtually applicable to any molecule.

The Direct Count Method. A drawback of the symmetry number method is that apparently it does not consider the number of microspecies constituting a given macrospecies. Thus it was felt that a systematic method to count the number of microspecies would be very useful to probe the scope of the symmetry number method. Upon searching the literature, we found that Bishop and Laidler (BL) proposed a method to count the number of microspecies for a generic equilibrium of the type shown in eq 8:[9]

$$A + B \underset{r}{\overset{l}{\rightleftharpoons}} C + D \tag{8}$$

According to BL, the statistical factor for such an equilibrium is simply given by the ratio l/r, where l is the number of chemically plausible different sets (microspecies) of C + D that can be formed if all identical atoms in A and B are labeled, and r is similarly defined as the number of chemically plausible different sets of A + B that can be formed if all identical atoms in C and D are labeled. BL also proved that, for the generic equilibrium in eq 8, the ratio l/r is equal to the ratio of symmetry numbers, $\sigma_A\sigma_B/\sigma_C\sigma_D$.

In order to illustrate the application of the BL method, consider the hypothetical equilibrium shown in eq 9:[18]



(9)

By labeling a linear alkane, it is easy to show that the hydrogens of the terminal methyl groups can be combined in nine different ways to form the corresponding cycloalkane and H_2. Conversely, by labeling the products, it can be seen that, for each of the n equivalent C−C bonds of the ring, there are two different orientations for the addition of H_2. The statistical factor of the reaction is thus $9/2n$. The same result could be obtained by the symmetry number method, by considering σ (n-alkane) = 18, σ (n-cycloalkane) = $2n$, and $\sigma(H_2) = 2$.

To find application in self-assembly, the method of BL needs to be generalized to the equilibrium shown in eq 1, in which a given species may appear with a stoichiometric coefficient greater than 1. Let us consider a reaction in which a molecules of the reactant A are needed to form the products. After having distinctly labeled all the identical atoms of all the A molecules (for example, if the reaction requires three methane molecules, these molecules should be labeled as $C^1H^1H^2H^3H^4$, $C^2H^5H^6H^7H^8$, and $C^3H^9H^{10}H^{11}H^{12}$, with the *label*-chirality of the three molecules given by the ordered sequence of hydrogens being

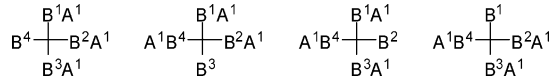


Figure 1. Total microspecies of the product obtained by binding three identically labeled molecules A to the tetravalent receptor B.

the same), let us form with the atoms of these molecules the whole set of plausible microspecies of the products. The number of generated microspecies, however, is redundant because, for example, two of them could be converted into each other by interchanging the labels of the atoms belonging to, say, the first A molecule with those of the corresponding atoms belonging to, say, the second A molecules. These two microspecies would actually be identical because the two A molecules are indistinguishable. Since in total $a!$ permutations of this type are possible, the number of distinct microspecies is obtained by dividing the number of generated microspecies by $a!$. Generalizing this result to the case of eq 1, $l =$ (number of generated microspecies of products)/$a!b!$ and $r =$ (number of generated microspecies of reactants)/$c!d!$.

An interesting case to analyze with this procedure is the binding of i molecules of a monovalent ligand, A, to an n-valent receptor, B, as in the case of the binding of O_2 (a monovalent ligand) to hemoglobin (a tetravalent receptor). Let us consider the sequential addition of i molecules of A, distinctly labeled as A^1, A^2, ..., A^i, to an n-valent receptor, B, whose identical binding sites are labeled B^1, B^2, ..., B^n. A^1 can be placed in any of the n distinct sites of the receptor B, A^2 can be placed in one of the $n - 1$ sites available, A^3 can be placed in one of the remaining $n - 2$ sites, and so on, up to the last ligand, A^i, which has $n - i + 1$ sites available. The total number of microspecies of the product that can be generated is thus given by the product $n(n - 1)(n - 2)\cdots\cdots(n - i + 1)$ equal to $n!/(n - i)!$. Division by $i!$ to correct for the redundancy of the generated microspecies gives the number of distinct microspecies of the product, l, corresponding to the number of combinations obtainable by putting i ligands on n distinct sites. The number of distinct microspecies of the reactants, r, is equal to 1, because there is only one way in which the reactants can be formed from the adduct. The result $r = 1$ is a general characteristic of all self-assembly equilibria starting from the constituent building blocks, because they are simple addition reactions. Accordingly, the statistical factor $\Omega_i = l/r$ is just given by the binomial coefficient (eq 10):

$$\Omega_i = \frac{n!}{(n - i)!i!} \tag{10}$$

A second approach, aimed at avoiding the redundancy of generated microspecies so that the final division by the factorial of stoichiometric coefficients is no longer required, can be envisaged. According to this approach, corresponding atoms of identical molecules are identically labeled; for example, if the reaction requires three methane molecules, these molecules should be identically labeled as $C^1H^1H^2H^3H^4$, $C^1H^1H^2H^3H^4$, and $C^1H^1H^2H^3H^4$. As an application, consider the binding of three identical ligand molecules A, all labeled as A^1, to a tetravalent receptor B. Here, the distinct microspecies shown in Figure 1 are generated, whose numbers correspond to that given by eq 10 with $i = 3$ and $n = 4$.

In the case of self-assembly processes, the second approach is simpler and will be used in the following. It is important to remark, however, that while the first approach, involving distinct labeling of identical molecules followed by division of the number of microspecies by the factorial of stoichiometric coefficients, is of general applicability, the second approach, involving identical labeling of identical molecules and the taking of the number of microspecies as such, is only valid if none of the generated labeled molecules presents symmetry axes. The following example illustrates the fact that the second approach does not always work. Consider the exchange reaction in eq 11:

$$A{-}A + B{-}B \rightleftharpoons 2\ A{-}B \tag{11}$$

It must be stated beforehand that the statistical factor of this process, obtained by the symmetry number method, is 4, because both of the reactants have a twofold symmetry axis.[19] Now, let us apply the direct count method by labeling the reactants as $A^1{-}A^2$ and $B^1{-}B^2$. If A^1 interacts with B^1, then A^2 is forced to interact with B^2, yielding the product set $A^1{-}B^1 + A^2{-}B^2$, whereas, if A^1 interacts with B^2, then A^2 is forced to interact with B^1, yielding the product set $A^1{-}B^2 + A^2{-}B^1$, thus two microspecies of the products are formed ($l = 2$). Labeling of the products according to the first approach yields ($A^1{-}B^1 + A^2{-}B^2$), which produces only one microspecies for the reactants, that is, $A^1{-}A^2 + B^1{-}B^2$. According to the first approach, to obtain the statistical factor r, we must divide the number of microspecies by the factorial of the stoichiometric coefficient that, in this case, is 2, so $r = 1/2$. Thus the statistical factor obtained by the first approach correctly predicts $l/r = 4$. Labeling of the products according to the second approach yields ($A^1{-}B^1 + A^1{-}B^1$), which produces 1 microspecies for the reactants, that is, $A^1{-}A^1 + B^1{-}B^1$; however, it appears that both labeled molecules present a twofold symmetry axis, thus the result obtained by the second approach, $r = 1$, must be rejected.

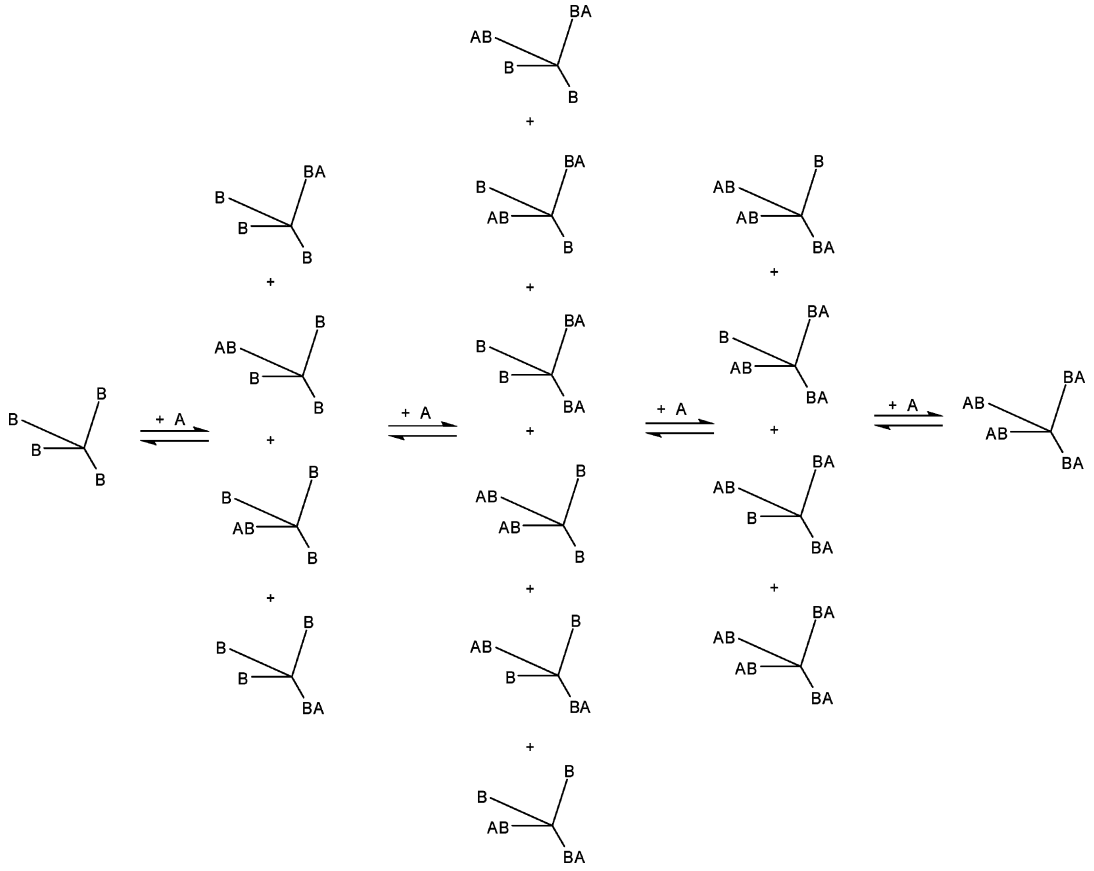
Discussion

More on the Binding of a Monovalent Ligand to a Multivalent Receptor. In the previous section we discussed the case of the binding of i molecules of a monovalent ligand, A, to an n-valent receptor, B, in the light of the direct count method. To treat this case, we made no assumptions about the geometry of the receptor. Although the evaluation of statistical factors by the symmetry number method is tied to the geometry of the species in equilibrium, we want to remark that, also in the light of the symmetry number method, *statistical factors are not affected by the geometry of an assembly, provided that only intermolecular bonds are involved in its formation.* To this end, we consider two limit geometries for a tetravalent receptor B: one asymmetric and the other tetrahedral. Sequential binding of the ligand A to the asymmetric receptor is illustrated in Scheme 1.

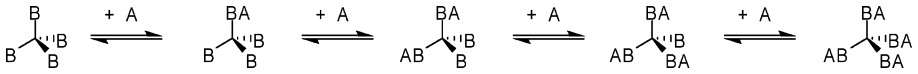
In this case, the symmetry number is 1 not only for the receptor but also for the intermediate adducts. Binding of the first molecule of A to the asymmetric receptor B produces a mixture of four distinct isomeric adducts BA, whose apparent symmetry number, according to eq 6, is 1/4. Thus, by eq 2, the statistical factor for the addition of one molecule of ligand A is 4. By adding two molecules of ligand A, a mixture of six distinct isomers BA_2 is formed, whose apparent symmetry number is 1/6, yielding a statistical factor of 6. Upon going on, it is easy to show that the overall series of statistical factors is 4, 6, 4, and 1, which is exactly the same as that obtainable by the binomial expression (eq 10) for i running from 1 to 4. Consider now the tetrahedral receptor (T_d) shown in Scheme 2; its symmetry number is 12.

By addition of one ligand molecule A, only one isomer BA is formed, whose symmetry number is 3 (C_{3v}). Thus, according

SCHEME 1



SCHEME 2



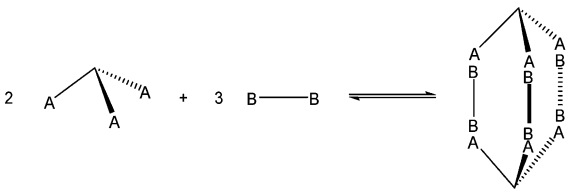
to eq 2, the statistical factor is 4. Upon the addition of two ligand molecules A, only one adduct BA_2 is formed, whose symmetry number is 2 (C_{2v}), and thus the statistical factor is 6. It is easy to show that, also in this case, the same series 4, 6, 4, and 1 is obtained, confirming that, when only intermolecular bonds are involved, the geometry of the receptor is of no concern.

It is interesting to note that the statistical factor for the binding of a monovalent ligand to a transition metal ion can be obtained by considering the metal ion as a multivalent receptor, provided that it has a well-defined number of coordinating ligands, B, undergoing substitution by a ligand A.[20] Indeed the statistical ratio of two successive stepwise stability constants in common textbook use[21] (eq 12) can be regarded as a consequence of eq 10, with Ω_i being the statistical factors related to the overall stability constants β_i:[20]

$$\frac{K_{i+1}}{K_i} = \frac{i(n-i)}{(i+1)(n-i+1)} \tag{12}$$

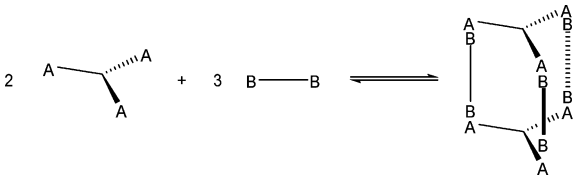
Note that eq 10 would not be correct for the addition of a monovalent ligand to a naked metal ion in the gas phase, because a naked metal ion has spherical symmetry (R_3, $\sigma = 1$) with no intrinsic coordination sites. In fact, the symmetry of a transition metal ion in solution depends on the geometry of its coordination sphere.[22]

Assemblies with Intramolecular Bonds. In contrast with the case above, *statistical factors for assemblies whose formation involves intramolecular bonds are strongly dependent on the geometrical features of the assembly as well as on those of its constituents*. Accordingly, for such assemblies, results of combinatorial approaches not taking geometry into account should be regarded with skepticism. To illustrate this point, consider the formation of the two assemblies in eqs 13 and 14, both having D_{3h} symmetry ($\sigma = 6$):



(13)

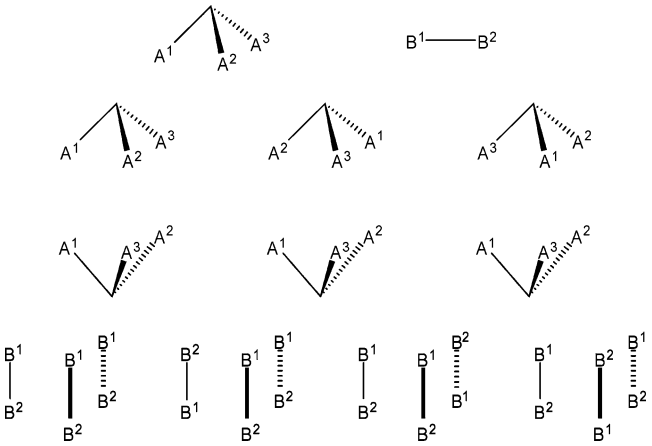
In the first case, the trivalent building block is pyramidal (C_{3v})



(14)

and does not undergo inversion, whereas, in the second case, it is planar (D_{3h}). By applying the symmetry number method, it is found that, for the first reaction, the statistical factor is $3^2 \times 2^3/6 = 12$, whereas, for the second one, it is $6^2 \times 2^3/6 = 48$. It is not at all intuitive why the change of the geometry of the

CHART 1



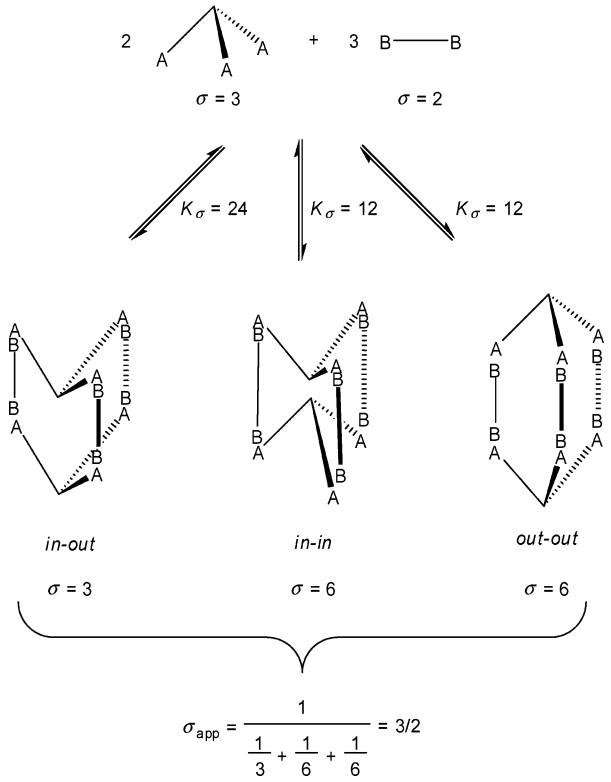
trivalent ligand produces such a change in the statistical factor. It is our opinion that difficulties of this type have unjustifiably hampered the widest acceptance of the symmetry number method in the self-assembly arena. In order to rationalize this result, let us apply the direct count method to eq 13. We follow the method of labeling the corresponding atoms of identical molecules with identical labels so as to avoid the final division of the number of microspecies by the factorial of stoichiometric coefficients. Labeling of the tripodal ligand produces a chiral arrangement of the A groups; we can say that the ligand is *label*-chiral. Indeed, in analogy with the well-known Cahn−Ingold−Prelog chirality rule,[23] for an observer looking at the pyramidal ligand from the base of the pyramid, the sequence $A^1-A^2-A^3$ can be either clockwise (*label-R*) or counterclockwise (*label-S*). The choice between the two orientations is arbitrary (we chose the *label-R* sequence shown in the first structure of Chart 1); the important point is that all the molecules of the ligand, being identical to each other and not undergoing inversion, are identically labeled.

It is evident that there are three distinct configurations for the two trivalent ligands constrained in the geometry of the product; these are obtainable by rotating one ligand with respect to the other, as shown in the second row of Chart 1. Note that the three configurations look identical if observed either from the top or from the bottom. As to the configurations of the divalent ligands, they must look different if observed from the top or from the bottom, otherwise, when mixed with the above configurations, would produce identical microspecies. Accordingly, there are only four configurations of the three divalent ligands constrained in the geometry of the product that satisfy this requirement; one has the three ligands aligned parallel, whereas the others have one of the ligands, in turn, that is aligned antiparallel to the other two ligands, as shown in the third row of Chart 1. Since the two sets of configurations are independent of each other, they can be mixed to form 12 distinct microspecies of the product, thus $l = 12$ and $r = 1$. The overall statistical factor is 12, in agreement with the symmetry number method.

In obtaining the above statistical factor, we have implicitly assumed that the *in*−*out* (C_{3v}) and *in*−*in* (D_{3h}) isomers, shown in Scheme 3, are significantly more strained than the *out*−*out* isomer, so that they do not form.

It is interesting to consider the case in which the *in*−*out* and *in*−*in* isomers have the same free energy, apart from the entropy of symmetry, as that of the *out*−*out* isomer. By the symmetry number method, it is easy to show that the statistical factors, K_σ, for the two isomers are 24 and 12, respectively. This

SCHEME 3

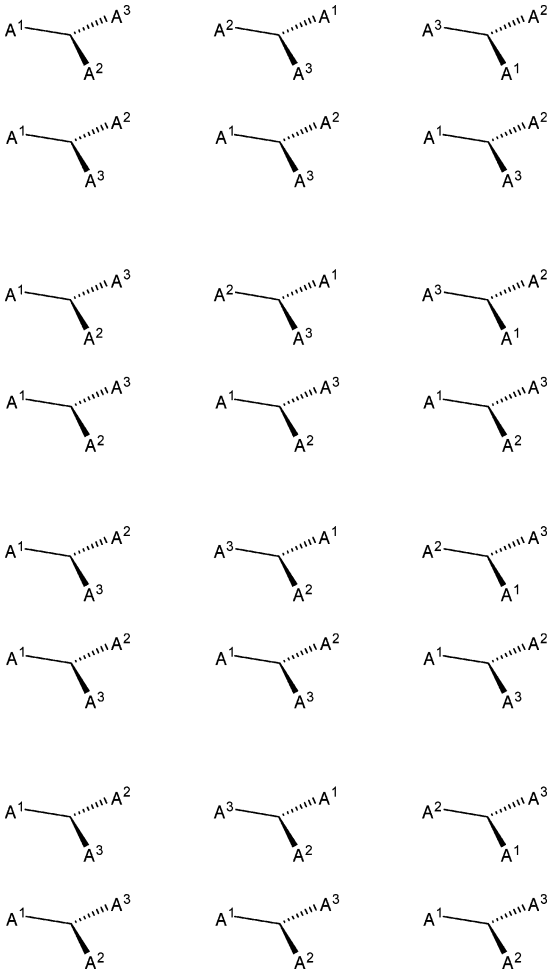


conclusion is readily confirmed by the direct count method. Indeed, the *in*−*out* isomer can be built by considering six different configurations for the two tripodal ligands: three are obtained by mutual rotation of the ligands in the arrangement in which the top ligand is *in* and the bottom ligand is *out*, and three are obtained the other way round, with the top ligand *out* and the bottom ligand *in*. Mixing of the six configurations with the four configurations of the divalent ligands shown in Chart 1 gives 24 distinct microspecies. For the *in*−*in* isomer, analogously to the *out*−*out* isomer, only 12 microspecies are possible. Considering the isomeric mixture as a whole, the overall statistical factor would be given by the sum of the individual statistical factors that is 48. The same result could be obtained by considering the apparent symmetry number of the mixture, $\sigma_{app} = 3/2$, given by eq 5.

As a further variation, consider the case in which the tripodal ligand can undergo fast inversion. The symmetry number of the tripodal ligand changes from 3 to 6 because of the double degeneracy of the potential well ($\sigma_{flux} = 2$), then the statistical factors for the *in*−*out*, *in*−*in*, and *out*−*out* isomers, assumed to be still distinguishable from each other,[13] become 96, 48, and 48, respectively. These figures are readily explained by the direct count method. Indeed, if we start with the *label-R* tripodal ligand, after a brief time, it becomes *label*-racemic because of the fast inversion process. Then to build the microspecies of the products, we have at disposal both the *label-R* and the *label-S* ligands. For the formation of the *out*−*out* isomer, 12 arrangements of the top and the bottom tripodal ligands can be generated, that is, three of each, due to mutual rotation, of *label-R*/*label-R* (shown in the second row of Chart 1), *label-R*/*label-S*, *label-S*/*label-R*, and *label-S*/*label-S*, that, when mixed with the four configurations of the divalent ligands, give 48 microspecies. The microspecies for the *in*−*out* and the *in*−*in* isomers are similarly found.

Now let us apply the direct count method to the self-assembly in eq 14. Owing to the planarity of the trivalent ligand, there is

CHART 2



only one way to label its binding groups. However, the two faces are not identical; they are *label*-enantiotopic. Indeed looking at the two faces, the sequence $A^1-A^2-A^3$ can be either clockwise (*label-Re* face) or counterclockwise (*label-Si* face).[24] Accordingly, for an observer placed inside the assembly, 12 arrangements of the top and the bottom trivalent ligands are possible, namely, three of each, due to mutual rotation of *label-Re*/*label-Re* faces (first row in Chart 2), *label-Re*/*label-Si* faces (second row in Chart 2), *label-Si*/*label-Re* faces (third row in Chart 2), and *label-Si*/*label-Si* faces (fourth row in Chart 2), that, when mixed with the four configurations of the divalent ligands, give 48 microspecies.

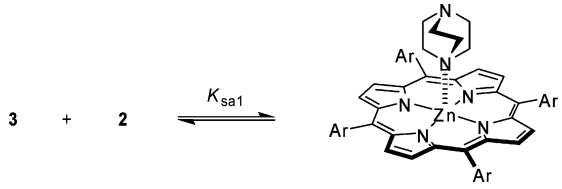
Other examples could be given; however, we feel that those above are sufficient to illustrate the elegance and power of the symmetry number method as well as the ability of the direct count method to rationalize the results provided by the former.

Application to a Real Case. It is useful to illustrate the application of both the symmetry number method and the direct count method to an interesting case that has been recently published, namely the 1,4-diazabicyclo[2.2.2]octane (DABCO)-induced self-assembly of a trisporphyrin double-decker cage.[3d] The building blocks are the tris-zincporphyrin **1** and DABCO (**2**) (Chart 3).

When mixed in a ratio of 2:3 in chloroform solution, they self-assemble to form a stable cage induced by the coordination of the nitrogen atoms of DABCO to zinc. As one of the authors has shown,[4] the self-assembly equilibrium constant, in the absence of cooperativity effects, can be factorized according to eq 15, where n is the number of assembly components, b is number of bonds joining them, and $K_{\text{inter(m)}}$ and $K_{\text{intra(m)}}$ are the microscopic inter- and intramolecular equilibrium constants, respectively.[25] Note that $n - 1$ represents the number of intermolecular bonds, and $b - n + 1$ represents the number of intramolecular bonds required to form the assembly.

$$K_{\text{sa}} = K_{\sigma} K_{\text{inter(m)}}^{n-1} K_{\text{intra(m)}}^{b-n+1} \tag{15}$$

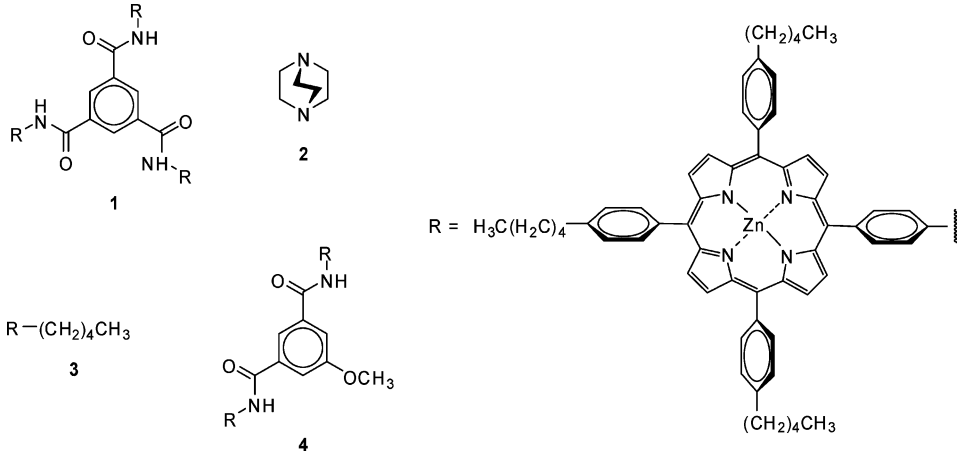
Before discussing the statistical factor of the cage, it is useful to examine how the values of $K_{\text{inter(m)}}$ and $K_{\text{intra(m)}}$ can be obtained. The first constant can be obtained by studying the equilibrium between zinc-5,10,15,20-tetrakis(4-pentylphenyl)-porphyrin (**3**) and DABCO (**2**) reported in eq 16:



(16)

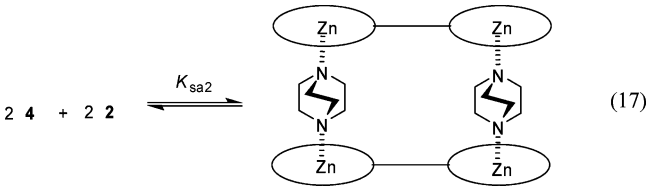
For the equilibrium in eq 16, $n = 2$ and $b = 1$, thus, according to eq 15, the microscopic intermolecular constant, $K_{\text{inter(m)}}$, is obtained by correcting the observed constant K_{sa1} for the statistical factors inherent to eq 16 ($K_{\text{inter(m)}} = K_{\text{sa1}}/K_{\sigma 1}$). The symmetry numbers of **3** (D_{4h}) and **2** (D_{3h}) are 8 and 6, respectively, while the symmetry number of the adduct is 12, because $\sigma_{\text{ext}} = 1$ and $\sigma_{\text{int}} = 3 \times 4$, corresponding to the product of the two rotation axes of the rotors (porphyrin = 4 and DABCO = 3),[26] thus, according to eq 2, $K_{\sigma 1} = 4$. The statistical

CHART 3



factor 4 can be also obtained by the direct count method considering that DABCO has two equivalent nitrogen atoms and that the two faces of the planar zincporphyrin ring are *label*-enantiotopic. From the experimental value log $K_{sa1} = 5.11 \pm 0.03$,[3d] the value log $K_{inter(m)} = 4.51 \pm 0.03$ is thus obtained.

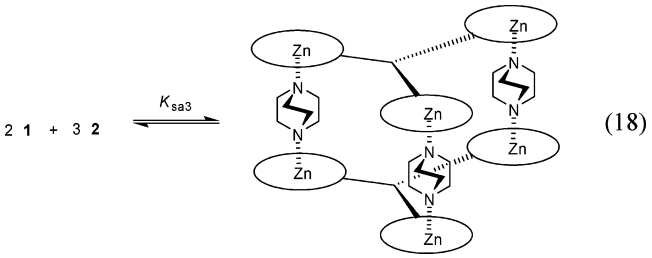
The constant $K_{intra(m)}$ can be obtained by studying the equilibrium between the bis-zincporphyrin (**4**) and DABCO (**2**), reported schematically in eq 17:



(17)

For the equilibrium in eq 17, $n = 4$ and $b = 4$, thus, according to eq 15, the microscopic intramolecular constant, $K_{intra(m)} = K_{sa2}/(K_{inter(m)}^3 K_{\sigma 2})$, where $K_{\sigma 2}$ is the statistical factor inherent to eq 17. The symmetry number of **4** is 8 because of the twofold symmetry axis, and $\sigma_{int} = 2^2$ for the internal rotation of each of the two terminal porphyrin rings; the symmetry number of **2** is 6, while that of the cyclic adduct is 36, because of two independent external twofold axes and $\sigma_{int} = 3^2$ for the internal rotation of the two DABCO subunits; accordingly, $K_{\sigma 2} = 64$. This figure can also be obtained by the direct count method considering that there are two distinct arrangements for the two labeled DABCO subunits (parallel and antiparallel), as well as two distinct arrangements (parallel and antiparallel) for the two bis-zincporphyrins; moreover, each of the four porphyrin rings in the cyclic adduct can bind DABCO with one or the other of the two *label*-enantiotopic faces. From the experimental value log $K_{sa2} = 18.8 \pm 0.1$,[3d] considering $K_{\sigma 2}$ and the value of $K_{inter(m)}$ above, log $K_{intra(m)} = 3.4 \pm 0.2$ is easily obtained.

The formation of the double-decker cage assembly can be schematized by eq 18:



(18)

For the equilibrium in eq 18, $n = 5$ and $b = 6$, thus, according to eq 15, $K_{sa3} = K_{\sigma 3} K_{inter(m)}^4 K_{intra(m)}^2$. The symmetry number of **1** is 48 because $\sigma_{ext} = 6$ and $\sigma_{int} = 2^3$ for the internal rotation of the three terminal porphyrin rings; the symmetry number of **2** is 6, while that of the cage adduct is 162, because of the D_{3h} external symmetry($\sigma_{ext} = 6$) and $\sigma_{int} = 3^3$ for the internal rotation of the three DABCO subunits; thus $K_{\sigma 3} = 3072$. This figure can be also obtained by the direct count method, considering that the cage assembly is reminiscent of the assembly shown in eq 14; in analogy with the discussion above, there are four distinct arrangements for the three labeled DABCO subunits, schematized in the second row of Chart 1, as well as 12 distinct arrangements for the two tris-zincporphyrins schematized in Chart 2; moreover, each of the six porphyrin rings in the cage assembly can bind DABCO with one or the other of the two *label*-enantiotopic faces. The estimated value of log K_{sa3} in the absence of cooperative effects is thus 28.4 ± 0.6, which is equal, within the estimated error, to the experimental log $K_{sa3} = 27.9 \pm 0.1$.[3d] According to the previously established criteria for the assessment of cooperativity in self-assembly processes,[4] it can be concluded that the self-assembly of the cage is a noncooperative process.

Conclusion

The symmetry number method is an easy and fast approach for the evaluation of statistical factors. In spite of its simplicity, it has not gained wide acceptance in the self-assembly arena, probably because its results may be counterintuitive. The direct count method presented here not only strengthens the role played by the symmetry number method, but also serves as a rationale for the apparently odd results the latter sometimes provides. The two methods are thus complementary to each other. The above examples point to the importance and subtlety of the role played by the geometrical features of assemblies involving intramolecular bonds.

References and Notes

(1) (a) Mammen, M.; Choi, S. K.; Whitesides, G. M. *Angew. Chem., Int. Ed.* **1998**, *37*, 2755−2794. (b) Lundquist, J. J.; Toone, E. J. *Chem. Rev.* **2002**, *102*, 555−578. (c) Badjic, J. D.; Nelson, A.; Cantrill, S. J.; Turnbull, W. B.; Stoddart, J. F. *Acc. Chem. Res.* **2005**, *38*, 723−732.

(2) (a) Lehn, J. -M. *Supramolecular Chemistry: Concepts and Perspectives*; VCH Publishers: Weinheim, Germany, 1995. (b) Philp, D.; Stoddart, J. F. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1154−1196. (c) *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. (d) Lindoy, L. F.; Atkinson, I. *Self-Assembly in Supramolecular Systems (Monographs in Supramolecular Chemistry)*; Stoddart, J. F., Ed.; The Royal Society of Chemistry: Cambridge, U.K., 2000. (e) *Science* (special section on Supramolecular Chemistry and Self-Assembly) **2002**, *295*, 2395−2421.

(3) (a) Hunter, C. A.; Tregonning, R. *Tetrahedron* **2002**, *58*, 691−697. (b) Baldini, L.; Ballester, P.; Casnati, A.; Gomila, R. M.; Hunter, C. A.; Sansone, F.; Ungaro, R. *J. Am. Chem. Soc.* **2003**, *125*, 14181−14189. (c) Ballester, P.; Costa, A.; Castilla, A. M.; Deyà, P. M.; Frontera, A.; Gomila, R. M.; Hunter, C. A. *Chem.−Eur. J.* **2005**, *11*, 2196−2206. (d) Ballester, P.; Oliva, A. I.; Costa, A.; Deyà, P. M.; Frontera, A.; Gomila, R. M.; Hunter, C. A. *J. Am. Chem. Soc.* **2006**, *128*, 5560−5569.

(4) (a) Ercolani, G. *J. Am. Chem. Soc.* **2003**, *125*, 16097−16103. (b) Ercolani, G. *Struct. Bonding* **2006**, *121*, 167−215.

(5) Kitov, P. I.; Bundle, D. R. *J. Am. Chem. Soc.* **2003**, *125*, 16271−16284.

(6) (a) Hamacek, J.; Borkovec, M.; Piguet, C. *Chem.−Eur. J.* **2005**, *11*, 5217−5226. (b) Hamacek, J.; Borkovec, M.; Piguet, C. *Chem.−Eur. J.* **2005**, *11*, 5227−5237. (c) Hamacek, J.; Borkovec, M.; Piguet, C. *Dalton Trans.* **2006**, 1473−1490.

(7) Benson, S. W. *J. Am. Chem. Soc.* **1958**, *80*, 5151−5154.

(8) (a) Benson, S. W. *Thermochemical Kinetics*, 2nd ed.; Wiley-Interscience: New York, 1976; pp 37−39. (b) Bailey, W. F.; Monahan, A. S. *J. Chem. Educ.* **1978**, *55*, 489−493.

(9) Bishop, D. M.; Laidler, K. *J. Chem. Phys.* **1965**, *42*, 1688−1691.

(10) DeTar, D. F. *J. Phys. Chem. A* **1998**, *102*, 5128−5141.

(11) Ercolani, G. *J. Chem. Educ.* **2005**, *82*, 1703−1708 and references therein.

(12) Pitzer, K. S. *J. Chem. Phys.* **1939**, *7*, 251−255.

(13) It is a common notion that two isomeric structures are considered distinguishable only if they are separated by an energy barrier significantly higher than *RT*. If the barrier separating two isomers is lower than *RT*, then they are in fact indistinguishable from each other, and the correct approach is to consider the symmetry number of the most symmetric among the energetically indistinguishable structures (see the discussion about the symmetry number of linear alkanes).

(14) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; Wiley: Chichester, U.K., 2004; pp 377−378.

(15) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley-Interscience: New York, 1969.

(16) This view corresponds to the model that polymer chemists call "the freely rotating chain".[15]

(17) In the case of cyclobutane, in which there is a single puckered conformation (D_{2d}) undergoing fluxional motion, the same result can be obtained by considering the product of $\sigma_{ext} = 4$ and $\sigma_{flux} = 2$.

(18) Mandolini, L. *Adv. Phys. Org. Chem.* **1986**, *22*, 1−111.

(19) To avoid circular arguments that justify the direct count method on the basis of the symmetry number method and vice versa, we wish to remark that the statistical factor 4 for the equilibrium in eq 11 can be also

obtained by the following reasoning. Let us envisage a stochastic process in which both the reactants and products are generated by sequential extraction of a couple of atoms from a large pool of A and B atoms present in equal amounts. The probability of extracting either an A or a B atom is 0.5. The formation of A_2 requires two consecutive extractions of A with a total probability of 0.25. Analogously, the formation of B_2 requires two consecutive extractions of B with the same total probability of 0.25. The formation of AB can be accomplished either by extracting A and then B or by extracting B and then A; the two ways of forming AB both have a probability of 0.25, so the probability to form AB in two consecutive extractions is 0.5. The statistical equilibrium constant is thus given by $K = 0.5^2/(0.25 \times 0.25) = 4$.

(20) This analogy strictly holds when ligands A and B have the same symmetry properties in the free and bound states. In this case, the binomial coefficient, given by eq 10, corresponds to the statistical factor of the overall stability constant β_i. In the case where the symmetry properties are different, the binomial coefficient needs to be corrected by a factor c^i to account for such differences. For example, consider the substitution of a water molecule bound to a metal ion with an ammonia molecule. In the reactants, bound water has $\sigma_{int} = 2$ and free ammonia has $\sigma = 6$ ($\sigma = \sigma_{ext}\sigma_{flux}$); in the products, free water has $\sigma = 2$ and bound ammonia has $\sigma_{int} = 3$; accordingly, $c = 2 \times 6/(2 \times 3) = 2$. The factor c is different from 1 because of the different symmetry properties of free and bound ammonia. Note that eq 12, being equal to $(\beta_{i+1}/\beta_i)/(\beta_i/\beta_{i-1})$ is not affected by the presence of this correction factor.

(21) Cotton, F. A.; Wilkinson, G. *Advanced Inorganic Chemistry*, 5th ed.; Wiley: New York, 1988; pp 41−45.

(22) In a previous work of one of us (ref 4a), analyzing the possible cooperativity in the formation of the famous Lehn's double-stranded trimetallic helicate in acetonitrile solution, the symmetry number of Cu^+ ions was taken to be equal to 1 as if it were a naked ion. Indeed it is known that Cu^+ in acetonitrile forms a tetrahedral complex with four solvent molecules (ref 21, p 758), thus the symmetry number to consider for Cu^+ in CH_3CN should be that characteristic of the tetrahedral group, namely 12. It is important to point out, however, that the conclusion drawn in ref 4a, about the absence of cooperativity in the formation of the Lehn's helicate, is not affected at all by the changing of the symmetry number of Cu^+ from 1 to 12, although the values of the estimated microscopic equilibrium constants $K_{inter(m)}$ and $K_{intra(m)}$ should be reconsidered.

(23) Eliel, E. L.; Wilen, S. H.; Mander, L. N. *Stereochemistry of Organic Compounds*; Wiley-Interscience: New York, 1994.

(24) For concepts and nomenclature regarding prostereoisomerism, see ref 23, Chapter 8.

(25) Introducing the notion of microscopic effective molarity, EM_m, defined as the ratio of $K_{intra(m)}/K_{inter(m)}$, eq 15 may also be written as $K_{sa} = K_\sigma K_{inter(m)}{}^b EM_m{}^{b-n+1}$. For the concept of effective molarity and its applications, see refs 4b and 18.

(26) Note that this way of calculating σ_{int} for two coaxial symmetric rotors is only valid when the order of the symmetry axis of one of the rotors is not a multiple of the order of that of the other. Otherwise, σ_{int} is equal to the order of the more symmetric rotor only. This is because the order of the rotor of lower symmetry is taken into account by the axis of external rotation coinciding with the axis of internal rotation. For example, in the case of the adduct formed from zincporphyrin (order of the axis = 4) and pyridine (order of the axis = 2), $\sigma_{int} = 4$ and $\sigma_{ext} = 2$.