Estimating the Entropic Cost of Self-Assembly of Multiparticle Hydrogen-Bonded Aggregates Based on the Cyanuric Acid·Melamine Lattice

Mathai Mammen, Eugene I. Shakhnovich, John M. Deutch, and George M. Whitesides*

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

Received May 28, 1997

The entropic component of the free energy of assembly for multiparticle hydrogen-bonded aggregates is analyzed using a model based on balls connected by rigid rods or flexible strings. The entropy of assembly, ΔS, is partitioned into translational, rotational, vibrational, and conformational components. While previously reported theoretical treatments of rotational and vibrational entropies for assembly are adequate, treatments of *translational entropy in solution* and of *conformational entropy*—often the two largest components of ΔS—are not. This paper provides improved estimates and illustrates the methods used to obtain them. First, a model is described for translational entropy of molecules in solution (ΔS_{trans}(sol)); this model provides physically intuitive *corrections* for values of ΔS_{trans}(sol) that are based on the Sackur−Tetrode equation. This model is combined with one for rotational entropy to estimate the difference in entropy of assembly between a 4-particle aggregate and a 6-particle one. Second, an approximate analysis of a model based on balls connected by rods or strings gives an approximate estimate of the maximum contribution of conformational entropy to the difference in free energy of assembly of flexible and of rigid molecular assemblies. This analysis, although approximate, is easily applied by all types of chemists and biochemists; it serves as a guide to the design of stable molecular aggregates, and the qualitative arguments apply generally to any form of self-assembly.

We outline a method for estimating the entropy of association of multiparticle, hydrogen-bonded aggregates in organic solvent. The model we use in this method is closer to physics than it is to chemistry: molecular detail is reduced to either isolated balls or balls connected by rigid rods or flexible strings. Although greatly simplified relative to the complexity of complete molecular detail, the model has broad utility: we illustrate the application of this model (which we call the "ball-rod-and-string" or BRS) model by estimating the differences in the entropies of association of three assemblies based on the cyanuric acid (CA)−melamine (M) lattice (Figure 1).[1−5] All three assemblies share the common feature that they are composed of three CA moieties and three M moieties, interacting through 18 hydrogen bonds in a planar arrangement. They differ in the number, structure, and conformational flexibility of the groups that connect two or more of the melamine moieties covalently. The simplest assembly is CA_3M_3 (**1**); this aggregate is composed of three CA molecules and three M molecules. The aggregates hub$(M)_3$:3CA (**3**) and flex$(M)_3$:3CA (**2**) are composed of three CA molecules hydrogen-bonded to tris(melamine). In hub$(M)_3$:3CA, the groups that link the three melamines covalently have few degrees of torsional freedom; we classify these linking groups ("tethers") as "rigid". In flex$(M)_3$:3CA, the covalent tethers are composed of a large number of freely rotating bonds; we classify these tethers as "flexible".

In analyzing the stabilities of these types of aggregates, we start by assuming that hydrogen-bonding interactions dominate the enthalpy of interaction in all three assemblies and that the *enthalpies* of assembly for all three are the same (all three are composed of 18 hydrogen bonds). The observed differences in the stabilities, with this assumption, is due to differences in the *entropies* of assembly. The goal of this work is to create a working model for estimating the entropy of assembly for these multiparticle aggregates (and, by extension, for others). The paper is organized into two sections. In the first section, we describe our approach to evaluating the entropy of association of rigid particles (particles that have no torsional degrees of freedom). This section first reviews established methods for estimating the vibrational and rotational entropy and the classical method for estimating translational entropy. We then describe a method of estimating a substantial, physically intuitive correction to the value for translational entropy estimated by the Sackur−Tetrode equation for molecules in solution. We use these methods collectively to estimate the difference in the entropy of association of aggregates **1** and **3**. In the second section, we describe an approach to evaluating the entropy of association for conformationally flexible particles (particles for which a number of torsional degrees of freedom exist). This section introduces a simple and approximate, but conceptually useful, model for conformational entropy based on geometric shapes and their volumes. In this model, mol-

(1) Mathias, J. P.; Simanek, E. E.; Seto, C. T.; Whitesides, G. M. *Macromol. Symp.* **1994**, *77*, 157−166.
(2) Seto, C. T.; Whitesides, G. M. *J. Am. Chem. Soc.* **1990**, *112*, 6409−6411.
(3) Seto, C. T.; Whitesides, G. M. *J. Am. Chem. Soc.* **1993**, *115*, 905−916.
(4) Seto, C. T.; Whitesides, G. M. *J. Am. Chem. Soc.* **1993**, *115*, 1330−1340.
(5) Whitesides, G. M.; Simanek, E. E.; Mathias, J. P.; Seto, C. T.; Chin, D. C.; Mammen, M.; Gordon, D. M. *Acc. Chem. Res.* **1995**, *28*, 37−44.

S0022-3263(97)00944-4 CCC: $15.00
Published on Web 05/28/1998

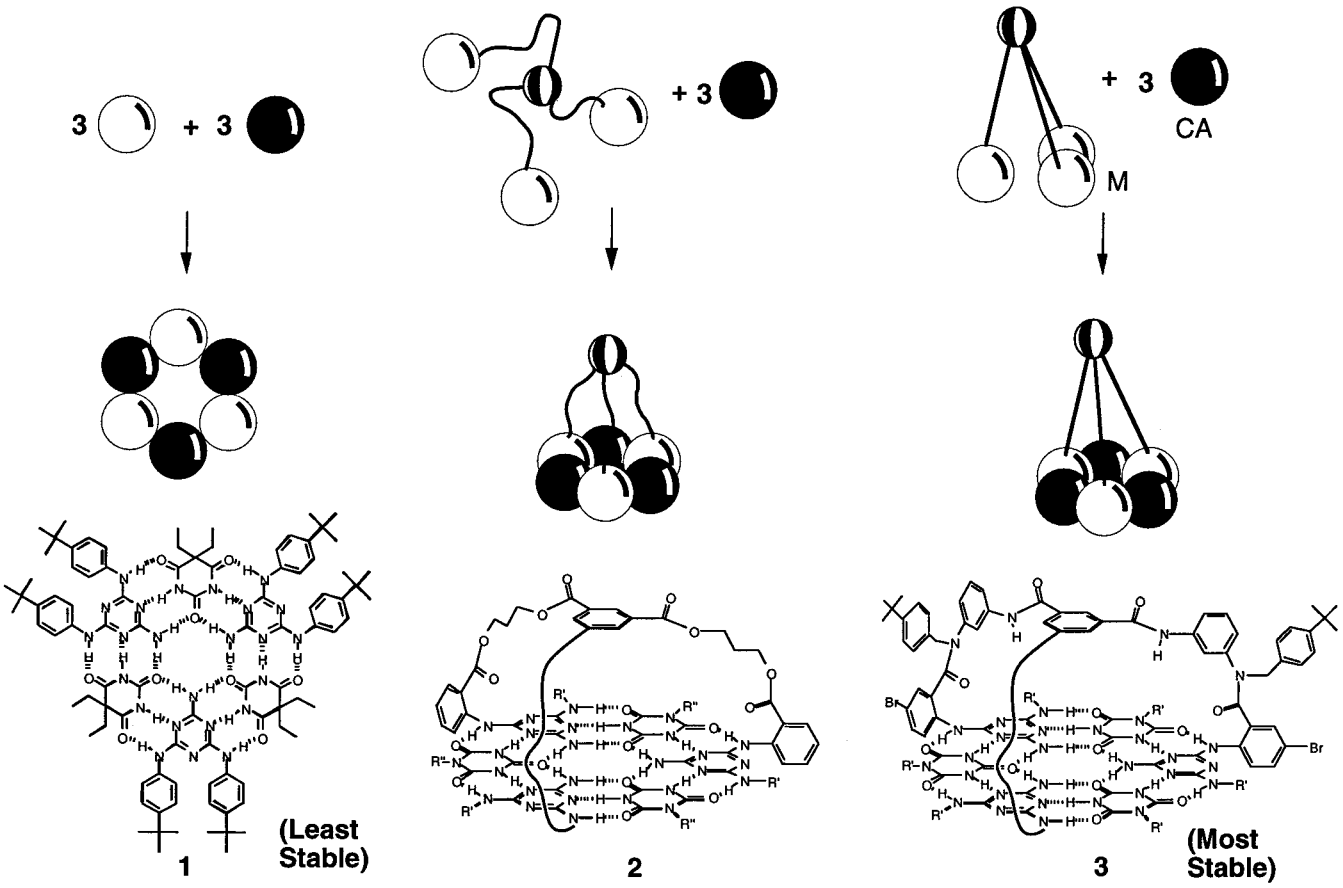

Figure 1. Molecular structures and simplified ball attached to rods and strings (BRS) schematic representatives of the aggregates 3CA:3M (**1**), flex(M)$_3$:3CA (**2**), and hub(M)$_3$:3CA (**3**) and of their formation from their constituent pieces. See refs 1−5 for for information regarding the synthesis and characterization of these aggregates.

ecules are represented at the level of balls attached to strings (flexible) or rods (rigid). We use this model to estimate the difference in the entropy of association of aggregates **2** and **3**.

Entropy of Association for an Aggregate of Rigid Particles

At the present time, it is impractical to estimate the entropy of association for structurally complex multiparticle assemblies using computer simulations: the number of translational and rotational degree of freedom is too high to sample reliably.[6] Absolute values for ΔG of assembly cannot, therefore, be predicted reliably using any computational package.[6] Instead, we use a theoretical model that estimates the changes in entropy that accompany aggregation by partioning the total entropy into four components and evaluating each component separately.

We partition entropy into translational, rotational, vibrational, and conformational components (eq 1). For

$$\Delta S = \Delta S_{\text{trans}} + \Delta S_{\text{rot}} + \Delta S_{\text{vib}} + \Delta S_{\text{conf}} \qquad (1)$$

rigid particles, ΔS_{conf} is, by definition, zero. Molecular and physical characteristics affect the translational, rotational, and vibrational entropies of rigid particles. Figure 2 suggests the relative magnitudes of each of the components of entropy for multiparticle, self-assembling systems of molecules and of the approximate uncertainties in each of these values based on current theoretical models.[7,8] The value of the translational entropy of assembly is always less than the value estimated using traditional methods, as we will dicuss in detail later.

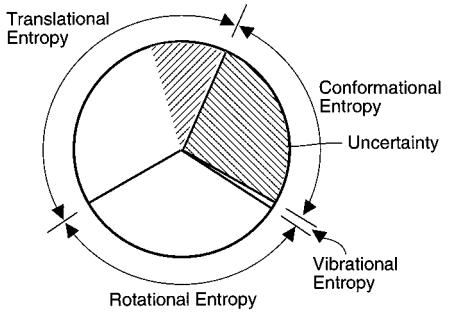

Figure 2. Approximate contributions of translation, rotational, vibrational, and conformational components to the total entropy of assembly. The uncertainties in each of these values are indicated by the hash marks, and their magnitudes are justified in the text. The chart indicates that the value of translational entropy is always less than the estimate (discussed in detail in the text) by some unknown, but large (~20−40%) amount. The uncertainty in the estimated value of the conformational entropy is as large as the estimate itself.

(6) Edholm, O.; Berendsen, J. C. *Mol. Phys.* **1984**, *51*, 1011−1028.

(7) Steinberg, I. Z.; Scheraga, H. A. *J. Biol. Chem.* **1963**, *238*, 172−181.

(8) Williams, D. H.; Cox, J. P.; Doig, A. J.; Gardner, M.; Gerhard, U.; Kaye, P. T.; Lai, A. R.; Nicholls, I. A.; Salter, C. J.; Mitchell, R. C. *J. Am. Chem. Soc.* **1991**, *113*, 7020−7030.

Table 1. Comparison of the Absolute Values of Experimental Entropies of Monoatomics at 1 atm, T = 298 K, to Those Values of Calculated Using Eq 2 (the Sackur−Tetrode Equation) or Using the Free Volume Model (See Text for Details of This Calculation)[a]

			Sackur−Tetrode model		free volume model	
atom	standard state	exp S (J mol^{-1} K^{-1})[b]	calc S (J mol^{-1} K^{-1})	ΔS (J mol^{-1} K^{-1})[c]	calc S (J mol^{-1} K^{-1})	ΔS (J mol^{-1} K^{-1})
He	gas	126.06	125.98	0.92	125.98	0.92
Ne		146.23	146.23	0.00	146.23	0.00
Ar		154.72	154.68	−0.04	154.68	−0.04
Kr		163.97	163.93	−0.04	163.93	−0.04
Xe		169.58	169.54	−0.04	169.54	−0.04
Li		138.6	137.9	−0.7	137.9	−0.7
Na		153.64	153.55	−0.09	153.55	−0.09
K		160.2	159.5	−0.7	159.5	−0.7
Rb		170.0	169.2	−0.8	169.2	−0.8
Al		164.43	164.68	0.25	164.68	0.25
Ag		172.88	172.84	−0.04	172.84	−0.04
Hg		174.89	174.81	−0.08	174.81	−0.08
He	aq[d]	55.6	99.6	44.0	40.0	−15.6
Ne		66.1	119.8	53.7	60.3	−5.8
Ar		59.4	128.4	69.0	68.9	9.5
Kr		61.5	137.7	76.2	78.2	16.7
Xe		65.7	143.3	77.5	83.8	18.1
Hg	liquid	77.4	113.3	35.9	35.9	0.0

[a] Experimental values are taken from the *CRC Handbook of Chemistry and Physics*, 70th Ed., 1990 (pp D15−65). [b] Exp S = experimental entropy. [c] ΔS = experimental value minus calculated value (calc S). [d] 1 mol/L in water.

Since no good model exists for conformational entropy, the uncertainty in the estimated value is as large as the estimate. The following sections justify the assertions in this figure.

Translational Entropy. The magnitude of the translational entropy, S_{trans}, reflects the possible number of unique arrangements of a collection of molecules in a given space (Sackur−Tetrode equation; eq 2);[9] quantitatively, S_{trans} is related to the logarithm of that number. The value of S_{trans} correlates positively with temperature (T, K) and mass (M, g/mol) of the particle and correlates inversely with the concentration [X] (mol/L) of the particles. The first numerical term in eq 2, $10^{-15/2}$, converts units of m^3 and kg into L and g, respectively. Equation 2 also includes Plank's constant, h (J s^2), the Boltzmann constant, k (J mol^{-1} K^{-1}), the fundamental constant e (unitless), and Avogadro's number N_A (unitless). The expression within the brackets is unitless. The value of the constant term in the numerical evaluation of eq 2 depends on the choice of standard state and therefore on the value of [X]: as a gas, the standard state is [X] = 4.46 × 10^{-2} mol/L (1 mol at 1 atm occupies 22.4 L) and the constant is 36.9 J mol^{-1} K^{-1} as shown; as a pure liquid, the value of [X] depends on the density of the liquid at 25 °C and the value of the constant is different for different substances (for typical organic solvents, $[X]_{liquid} \approx 10$ mol/L, and the value of the constant is −8.0); and as a solution, the standard state is [X] = 1 mol/L and the constant is 11.1 J mol^{-1} K^{-1}; Equation 2 is a classical (i.e. nonquantum mechanical) expression and is not valid at low temperatures (T < 10 K) where quantum effects on translational entropy are often important.

$$S_{trans} = R \ln\left[\left(\frac{10^{-15/2}}{N_A^4[X]}\right)\left(\frac{2\pi MRTe^{5/3}}{h^2}\right)^{3/2}\right] = 36.9 + 12.5 \ln M + 12.5 \ln T \quad (2)$$

Equation 2 successfully predicts the entropy of monatomic gases—that is, particles whose entropy consists solely of translational entropy—at 1 atm at 298 K (Table 1). The translational entropy of a molecule in the liquid or solution phase is, however, *significantly* lower than that predicted using eq 2 (Table 1 gives 7 examples).[10−12] Qualitatively, the principal reason for the failure of eq 2 to predict the translational entropy of a liquid accurately is that it ignores molecular volume, V_{molec} (defined as that enclosed by the van der Waals surface of the molecule); in a liquid, V_{molec} is a significant component of the total volume, V_{total}. We discuss this discrepancy quantitatively later.

In summary, there is still a substantial uncertainty regarding the value of translational entropy in solution. We later introduce a model that predicts translational entropy in solution more accurately than does eq 2 by using an extension of previously described *free volume* theory.

Rotational Entropy. The magnitude of the rotational entropy, S_{rot}, of a particle is related quantitatively to the logarithm of the number of unique rotational positions that are open to it (eq 3a).[13] The three principal moments of inertia for the molecule are I_A, I_B, and I_C (kg m^2). As in eq 2, the expression within the brackets of eq 3 is unitless. We define the molecular density of a

$$S_{rot} = R \ln\left[\pi^{1/2}\left(\frac{8\pi^2 RTe}{h^2 N_0}\right)^{3/2} (A_A I_B I_C)^{1/2}\right] \quad (3a)$$

(9) Gurney, R. W. *Introduction to Statistical Mechanics*, 1st ed.; McGraw-Hill Book Company: New York, 1949.

(10) Fujiwara, H.; Ohtaku, I.; Takagi, T.; Murata, S.; Sasaki, Y. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 1853−1856.

(11) Perlman, M. L.; Rollefson, G. K. *J. Chem. Phys.* **1941**, *9*, 362−369.

(12) A value of R ln 2 was added to the value predicted by eq 2 for all alkali metals. This correction accounts for the 2-fold degenracy of the lowest *electronic* state in these metals. A discussion of this effect can be found in Chapter 4 of Gurney, R. W. *Introduction to Statistical Mechanics*; McGraw-Hill: New York, 1949.

(13) The spacing between rotational energy levels is sufficiently small ($\Delta E_{rot} \ll kT$) at temperatures greater than ~10 K that this classical expression is accurate, and quantization of the energy levels can be ignored.

molecule as its molecular mass (kg/molecule) divided by its value of V_{molec} (m^3). The molecular density of organic molecules composed primarily of C, N, and O—for example, cholesterol, glucose, benzoic acid, insulin, M, and CA—is approximately constant: from this constant value, we relate the mass of a molecule to the moment of inertia of a sphere of equal mass (in which case $I_{sphere} = I_A = I_B = I_C$) (eq 3b).[14] We combine eqs 3a and 3b to give 3c and evaluate its constants. The value of rotational entropy

$$I_{sphere} = 8.3 \times 10^{-49}\ M^2 \qquad (3b)$$

$$S_{tot} = R \ln\left[\pi^{1/2}\left(\frac{8\pi^2 RTe}{h^2 N_0}\right)^{3/2} I_{sphere}^{3/2}\right] = 25.0 \ln M + 12.5 \ln T - 74.1 \qquad (3c)$$

in solution is, in most cases, the same as that in the gas phase, and the uncertainty in its magnitude is low (typically, the estimate is within 2% of the experimental value).[15,16]

Vibrational Entropy. The magnitude of the vibrational entropy, S_{vib}, of a molecule reflects the characteristic frequencies, ν_0 (cm^{-1}) of all its vibrational motions. High frequency (>100 cm^{-1}) motions, including bond stretches, bends, and wags, are modeled well as springs, with a characteristic spring constant; for high-frequency motions, only the lowest vibrational state is occupied at room temperature and the vibrational energy at that state is $(1/2)h\nu_o$. The Einstein equation provides an exact relationship between S_{vib} and the value of ν_o (eq 4). The S_{vib} is a function of the values of the variables ν_o and T and the fundamental constants h, R, the speed of light c (cm/s), and k.

$$S_{vib} = R\left(\frac{hc\nu_o}{kT}\right)\left(\exp\left(\frac{hc\nu_o}{kT}\right) - 1\right)^{-1} - R \ln\left(1 - \exp\left(-\frac{hc\nu_o}{kT}\right)\right) \qquad (4)$$

The total vibrational entropy of a molecule is the sum of all its S_{vib}, ΣS_{vib}. For the association of two molecules of A to form bimolecular complex A_2, the change in vibrational entropy ΔS_{vib} (J mol^{-1} K^{-1}) is defined as $\Delta S_{vib} = \Sigma S_{vib}(A_2) - 2\Sigma S_{vib}(A)$. Figure 3 illustrates that the magnitudes of TS_{vib} for almost all vibrations in a molecule at 25 °C are very small (typically much less than 2 kJ/mol). Furthermore, it is physically likely that the vast majority of high-frequency vibrations will be unaffected by assembly. We therefore assume in this work that the difference between small values is also small and that the value of ΔS_{vib} for molecular self-assembly is a negligible component of the overall change in entropy, and we do not consider it further.[17]

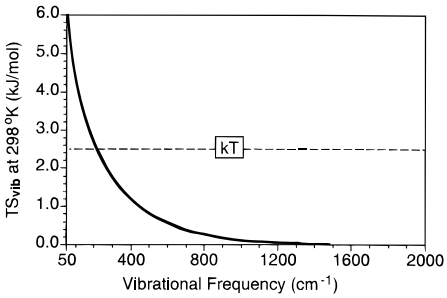

Figure 3. Value of vibrational entropy multiplied by temperature, TS_{vib} (J mol^{-1} K^{-1}) at 298 K as a function of vibrational frequency (cm^{-1}), as predicted using the Einstein equation (eq 4). The thermal energy at 298 K is ~2.5 kJ/mol.

A Better Model for Translational Entropy. Before describing our model, we further examine the assumptions of the Sackur—Tetrode equation (eq 2) in an attempt to understand why it yields grossly inaccurate estimates of translational entropy in condensed phases. The Sackur—Tetrode predicts a value for the translational entropy of a molecule dissolved in solution at 1 mol/L that is *larger* than the experimental value by a substantial amount (often >50% larger). Table 1 shows that for monatomics (species whose entropy consists solely of translational entropy) the Sackur—Tetrode equation is accurate in the gas phase but inaccurate in either liquid or solution phases. Several examples of such overprediction are given in Table 1 for atoms in solution whose entropy consists solely of translational entropy.[12] Since the change in translational entropy is the single largest component of the overall change in entropy during multiparticle molecular self-assembly, a large error in translational entropy of a single component results in a large error in the *change* in overall entropy for the assembly process. From a synthetic chemist's perspective, it is important to judge accurately the gain from covalently joining separate particles in order to reduce translational and rotational entropic costs of assembly, because such covalent connection comes at substantial synthetic effort.

The Sackur—Tetrode equation fails in a fundamental assumption; that is, molecules have no volume. This assumption is reasonable in the gas phase where the volume occupied by the molecules (that is, the molecule's own volume) is typically very small compared to the total volume. It is unreasonable in the condensed phase where most of the volume is occupied by the solvent.

We propose a method of correcting this overly large value that is centered on the idea of *free volume* in condensed phases;[18—20] free volume theory explicitly accounts for the volume occupied by the molecules. Our treatment differs from the free volume treatments proposed previously in one way: previous treatments attempted to use the free volume of a pure liquid (a solvent) to calculate the translational entropy of the molecules of solvent. We propose using the translational entropy of the solvent to estimate the translational entropy of a molecule dissolved in that solvent. This extension of free volume to estimation of translational entropies of molecules in solution is not rigorously correct because, in its simplest form, it assumes that the shapes and the sizes

(14) For molecules of regular composition (in our case, those that are composed predominatly of C, N, and O), we can assume a constant molecular density. The value of the moment of inertia, I (kg/m^2), for a solid sphere is $(2/5)MR^2$, where M is the mass (kg) and R is the radius of the sphere (m). The value of R for such a species of uniform density can be estimated from the value of M. For proteins, for example, $I = (8.306 \times 10^{-49})\ M^2$ (kg m^2), where M is in g/mol.

(15) Guggenheim, E. A. *Trans. Far. Soc.* **1941**, *37*, 97—105.

(16) Wertz, D. H. *J. Am. Chem. Soc.* **1980**, *102*, 5316.

(17) From previously published work on the CA·M rossette, we estimate that the vibrational frequency for umbrella motion (the most soft of the soft modes in this structure) is >50 cm^{-1} and therefore contributes negligibly to the value of ΔS_{ass} for aggregates **1**—**3**.

(18) Pierotti, R. A. *Chem. Rev.* **1975**, *78*, 717—726.

(19) Abraham, M. H. *J. Am. Chem. Soc.* **1981**, *103*, 6742.

(20) Frank, H. S. *J. Chem. Phys.* **1945**, *13*, 478—492.

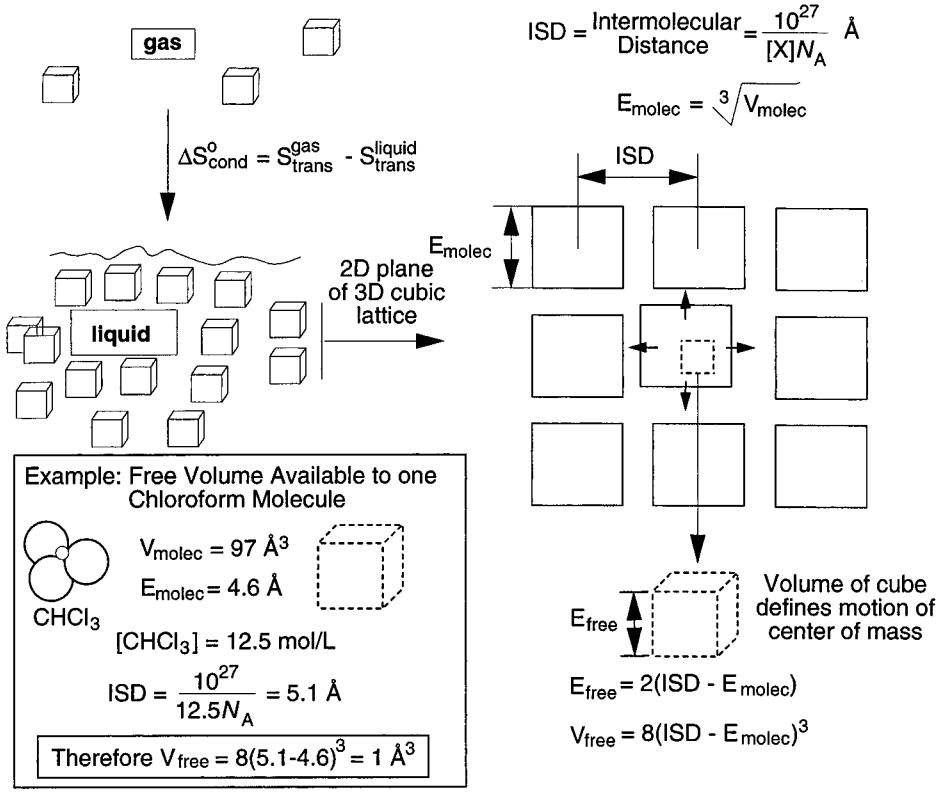

Figure 4. Overly large values of translational entropy determined using the Sackur−Tetrode equation (eq 2) are corrected using *free volumes* for pure liquids. Molecules of liquid of volume V_{molec} are represented by cubes of the same volume. Modeling the liquid as a regular cubic lattice, the spacing between the centers of the molecules are estimated from the concentration of molecules X, [X] (mol/L) in the liquid. The free volume available to a single molecule is estimated as the volume occupied by the center of mass moving in a cage defined by its nearest neighbors in the lattice. The free volume available to a single molecule of chloroform is 1 Å^3 and to 12.5 mol of chloroform (1 L) is 7.5 mL.

of the molecules of solute are the same as those of the molecules of solvent. We describe these limitations and assumptions in detail later. Nevertheless, it provides a *substantially* more accurate estimate of the translational entropy in solution than eq 2 and is therefore useful both numerically and conceptually to organic chemists interested in the design of multiparticle assemblies and to biochemists that study self-assemble (e.g. the spontaneous assembly of tobacco mosaic virus). We divide the remainder of this section into two parts. In the first part, we describe free volume theory as it is currently understood and apply it to the estimation of the translational entropy of molecules of solvent within a pure liquid. We describe how the accuracy of this model is tested in its ability to predict the experimental values for the entropy of condensation. In the second part, we extend this model to solutions and describe how this extended model can be used to estimate the translational entropies of molecules of interest dissolved in solution.

(a) Free Volume Theory. We have asserted that the volume open to a molecule in a liquid is substantially lower than the total volume. There are in principle two methods to calculate this lower volume in a liquid. One method accurately accounts for observed differences in the translational entropy in solution and the other method does not. We will first describe the method that does not because it is conceptually simpler and leads naturally into a discussion of the second method.

(i) Nonexcluded Volume, $V_{nonexcl}$. The volume available to the molecules of a pure liquid can be defined as the nonexcluded volume, $V_{nonexcl} = V_{total} - V_{molec}$. That is, this nonexcluded volume is that which is not occupied (or excluded) by the volume defined by van der Waals surfaces of the molecules. For typical organic solvents, the nonexcluded volume represents ~20−25% of the total volume. To put this number into perspective, the value of $V_{nonexcl}$ for organic crystals—substances in which translation (presumably) does not occur—is ~25% as well.[21] That is, since the unoccipied volume in both solids and liquids is approximately the same, we conclude that, typically, solids do not expand or contract significantly on melting. We will show later that the value for $V_{nonexcl}$ does not account accurately for the experimentally observed translational entropies in liquids. Since this model is inaccurate, we will not pursue it further and will next describe a more accurate model than it for estimating the available volume in a liquid.

(ii) Free Volume, V_{free}. The volume available to the molecules of a pure liquid can be defined by its characteristic free volume, V_{free}. The definition of free volume in statistical mechanics is *the volume occupied by the center of mass of one molecule of liquid moving randomly in a cage that is defined by its nearest neighbors* (Figure 4). This definition describes the volume of space open to a molecule of liquid in terms of the volume of space accessible to its center of mass. It is assumed in such a model that the liquid is described well by a regular array of hard cubes (or spheres, or other shapes), where the volume of each cube is equal to the molecular volume,

(21) The packing fraction in organic solids appears to vary much more widely than that for liquids and is <10% in one limit (diamond and graphite) and can be <65% (buckyballs).

V_{molec}. The spacing in this array of cubes is calculated using the concentration of molecules [X] (mol/L) in the liquid. Figure 4 illustrates schematically how V_{free} (Å^3) is estimated for any molecule from its values of [X] and V_{molec} (Å^3) (eq 5). The value of the constant C_{free} (unitless) in eq 5 varies slightly depending on the geometry chosen for both the regular array and the shape of the molecule.[20] For cubic molecules in a 3D cubic array, $C_{free} = 8$; for spherical molecules, $C_{free} = 6.3$.

$$V_{free} = C_{free}\left(\sqrt[3]{\left(\frac{10^{27}}{[X]N_o}\right)} - \sqrt[3]{V_{molec}}\right)^3; \qquad C_{free} = 8 \text{ for hard cubes} \quad (5)$$

As an illustrative example, we calculate the value of V_{free} for chloroform as follows: The density of chloroform is 12.5 mol/L, corresponding to one molecule of chloroform for every 134 Å^3 of space in the liquid. In a regular 3D cubic array, that corresponds to a spacing of approximately $(134)^{1/3} = 5.1$ Å between the centers of adjacent molecules of chloroform. In the hard cube approximation, the chloroform molecule is approximated as a cube with volume $V_{molec} = 97$ Å^3 (the value of V_{molec} can in general either be estimated simply using tabulated vdW radii for the atoms that comprise the solvent molecule or less simply, but more accurately, using available software packages such as Quanta). We used the latter method. This value of V_{molec} corresponds to a cube with each edge having length $E_{molec} = (97)^{1/3} = 4.6$ Å. The center of mass of a single cubic molecule is confined to a rigid cage defined by its nearest neighbors in the array. The motion of the center of mass defines a small cube, with each edge having length $E_{free} = 2 \times (5.1-4.6) = 1.0$ Å for chloroform, with volume $= (1.0)^3 = 1$ Å^3 (Figure 4). That is, the motion of the center of mass of a single molecule of chloroform occurs in volume of 1 Å^3 rather than in a volume of 134 Å^3. The *free* volume of a liquid is substantially less than the total volume that the liquid apparently occupies. The volume that should be used to calculate the density of the molecules for eq 2 should correctly correspond to this much lower volume. One liter of chloroform contains 12.5 mol. If the center of mass of each molecule of chloroform can occupy 1 Å^3 of volume, the centers of mass of 1 L of chloroform molecules then occupies $12.5N_A(1) \approx 7.5 \times 10^{24}$ Å^3 = 7.5 mL.

(b) Testing the Two Models Using Entropy of Condensation. The standard entropy of condensation, ΔS^o_{cond} is the change in entropy that occurs when 1 mol of molecules in the standard gaseous state (1 atm, 25 °C) condenses to 1 mol of molecules in the standard liquid state (1 atm, 25 °C). To a first approximation the changes in rotational and vibrational entropies are, in most cases, negligible during condensation.[10,16,18,19,22] The value of ΔS^o_{cond} therefore largely reflects the change in translational entropy during condensation.[20] Trouton's law states that the value of ΔS^o_{cond} for many organic liquids is nearly constant (-88 J mol^{-1} K^{-1}).[9] We interpret the physical basis of Trouton's law as simply a reduction in the volume available to the molecules in solution relative to the gas phase. If we assume first that the entire volume of liquid is available to the molecules (no excluded volume), then we can use eq 2 to calculate that the change in entropy due would be 44 J mol^{-1} K^{-1}; that is, the standard state of a gas is 1 mol in 22.4 L, where the standard state of a liquid is typically ~10 mol/L. This calculated value is grossly incorrect compared with the experimental value of ~88 J mol^{-1} K^{-1}: in other words, we cannot ignore the volume of the solvent molecules when using eq 2. Using the values of ΔS^o_{cond} as a benchmark, both models ($V_{nonexcl}$ and V_{free}) can be tested as two different ways of accounting for the volume available to the molecules in a liquid. We illustrate the difference between the two models again using chloroform as an example. Model I: using $V_{nonexcl} = 134$ Å^3 − 97 Å^3 = 37 Å^3, we use eq 2 to estimate that the value of ΔS^o_{cond} for chloroform is -55 J mol^{-1} K^{-1}. Model II: using $V_{free} = 1$ Å^3, we use eq 2 to estimate that the value of $\Delta S^o_{cond} = -85$ J mol^{-1} K^{-1}. We have similarly confirmed the greater accuracy of free volume theory over that of excluded volume for predictions of experimental values of ΔS^o_{cond} for a wide range of liquids.[23]

Even though the idea of free volume is established, with much of the groundbreaking work done by Eyring,[24] Frank,[20] and Kirkwood,[25] its contribution to an accurate model of the liquid state has been largely ignored in estimates of translational entropy by organic and physical organic chemists. We assert that the idea of free volume should play an important role in the estimation of translational entropy in liquid if properly used. On one hand, it is a pictorial idea, and on the other, it is capable of being brought into an exact relationship with thermodynamic measurement by allowing the numerical constant in eq 5 (C_{free}) to vary so that it accurately describes a large set of data.

(c) Extending Free Volume Theory to Solutions. We propose here a simple extension of free volume theory from liquids to solutions. That is, we propose that this model can be used to predict the translational entropy of any molecule dissolved in a liquid (and not just for the liquid itself) by calculating the value based on the Sackur−Tetrode equation using the apparent concentration of the molecule in the solvent (mol/L) and then subtracting a correction based on the free volume of the solvent. This extension is rigorously valid only for infinitely dilute solutions of solute. Two difficulties occur at high concentration of solute. First, the free volume of the solute begins to dominate the free volume of solution. Second, it is not trivial to estimate the free volume of large, irregularly shaped solutes (such as **1**, **2**, and **3**). We assume in this work that the concentrations of interest (≈ mmol/L) are sufficiently low that we can assume that the free volume of the solution is dominated by the free volume of the solvent. We illustrate the method with an example: 1 mol of methanol is dissolved in 1 L of chloroform such that [MeOH] = 1 mol/L. The free volume in chloroform (V_{free}) is much lower than its apparent volume (1 L). The value of V_{free} for 1 L of chloroform is only 7.5 mL (as shown earlier). Thus the effective concentration of methanol that *accurately re-*

(22) Ewing, G. *J. Chem. Phys.* **1969**, *73*, 168−174.

(23) We have examined benzene, toluene, methylene chloride, tetrahydrofuran, diethyl ether, and pyridine. For these solvents, the free volume theory is accurate (less than 10% error), and the excluded volume theory is not. The value of entropy of condensation for some protic solvents (acetic acid, methanol, ethanol) is predicted *reasonably* well (less than 20% error) by free volume theory but poorly by excluded volume. The value of the entropy of condensation for water is predicted poorly (greater than 30% error) by both methods.

(24) Eyring, H.; Hirschfelder, J. O. *J. Phys. Chem.* **1937**, *41*, 249−258.

(25) Kirkwood, J. G. *J. Chem. Phys.* **1950**, *18*, 380−382.

flects its translational freedom is not 1 mol per L but 1 mol per 7.5 mL = 134 mol/L. Using eq 2 without a correction, we calculate the translational entropy of 1 mol of methanol dissolved in 1 L of chloroform as S_{trans}(MeOH) = 125 J mol^{-1} K^{-1}. Using the value of V_{free} for chloroform, we calculate S_{trans}(MeOH) = 84.6 J mol^{-1} K^{-1}, a substantially lower number than that produced by eq 2. Interestingly, the difference between these two values is the difference between Δ^{o}_{cond} calculated using the Sackur–Tetrode equation and the experimental value of ΔS^{o}_{cond}. We will return to this connection in the next section. The effective concentrations (that is, based on the free volume rather than the apparent volume) can be much higher than the apparent concentration (e.g. 134 M compared with 1 M in the example with methanol). Analogously, the concentrations of species during intramolecular reactions can also be very high, and much higher than is apparently possible (up to 10^8 M). The important statement is that this free volume model, and the corresponding high concentrations in solution, brings the calculated values of ΔS^{o}_{cond} close to the observed thermodynamic values.

We propose two levels of correction to the estimate of translational entropy using eq 2. The zeroth-order correction is theoretical and uses eq 5 to predict V_{free} based on the calculated values of V_{molec} and the experimental values of $[X]_{liq}$. The next level of correction is more empirical, but more accurate, than the first and based completely on the experimental value of ΔS^{o}_{cond} for the solvent of interest.[26] Equations 6a and 6b are two modifications of eq 2 that use these two levels of correction. Equation 6a contains the experimental concentration of analyte, [analyte] (mol/L), and the free volume of the solvent $V^{solvent}_{free}$ (L). Equation 6b contains the experimental value for the entropy of condensation, $\Delta S^{solvent}_{cond,exp}$ (J mol^{-1} K^{-1}), and the value calculated by eq 2, $\Delta S^{solvent}_{cond,eq2}$ (J mol^{-1} K^{-1}). We use the more accurate eq 6b

$$S^{analyte}_{trans} = R \ln\left[\left(\frac{10^{-15/2}\ V^{solvent}_{free}}{N_o^4[analyte]}\right)\left(\frac{2\pi MRTe^{5/3}}{h^2}\right)^{3/2}\right] = 11.1 + 12.5\ln(T) + 12.5\ln(M) + 8.3\ln V^{solvent}_{free} \quad (6a)$$

$$S^{analyte}_{trans} = R \ln\left[\left(\frac{10^{-15/2}}{N_o^4[analyte]}\right)\left(\frac{2\pi MRTe^{5/3}}{h^2}\right)^{3/2}\right] + [\Delta S^{solvent}_{cond,exp} - \Delta S^{solvent}_{cond,eq2}] = 11.1 + 12.5\ln(T) + 12.5\ln(M) + [\delta S^{solvent}_{cond,exp} - \Delta S^{solvent}_{cond,eq2}] \quad (6b)$$

in this study. For chloroform, the value of S_{trans} using eq 6a is within 2% of the value calculated using eq 6b. This interpretation of free volume has not previously, to our knowledge, been used to correct the translational entropy of solutions of molecules.

In summary, the problems of predicting the entropy of association for rigid particles are in predicting its four components: translational, rotational, vibrational, and conformational entropies. The translational entropy is accurately predicted in the gas phase using the Sackur–Tetrode equation. The translational entropy in condensed phases (liquid, solution) can be estimated by using the value calculated in the gas phase (at the appropriate concentration) and then correcting this value using free volume theory and its extension. The rotational entropy for molecules in solution is predicted reasonably accurately using an existing model based on statistical mechanics. The change in vibrational entropy upon assembly is probably small, and we ignore it for this study. The change in conformational entropy for rigid particles is, by definition, zero. Figure 5 illustrates our model by predicting the difference in the entropy of association between aggregates **1** and **3**. As expected, the aggregate containing the larger number of particles (**1**) assembles at greated entropic cost. From NMR titration and competition experiments reported previously, we estimate that the difference in ΔG of assembly between **1** and **3** is approximately 100 kJ/mol, which is consistent with the theoretical estimate here.[27,28]

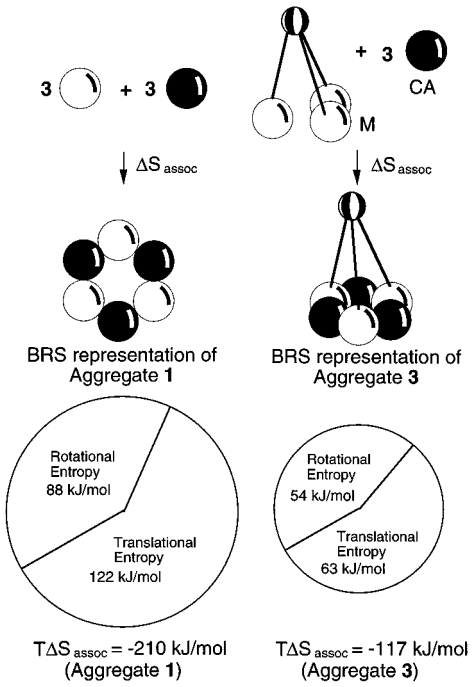

Figure 5. Difference between the entropies of association for the 6-particle aggregate **1** and the 4-particle aggregate **3**. These two aggregates and their components are approximated as balls attached by rods and strings. The changes in entropy upon assembly are calculated using eqs 1, 3c, and 6b.

Entropy of Association for an Aggregate of Flexible Particles

We classify molecular linkers that are comprised mostly (>75%) of single bonds as "flexible". In order to estimate the impact of conformational entropy of multiparticle assembly, we consider a simple model (the BRS model): the interacting (e.g. hydrogen bonding) molecules are balls; these balls may be free, attached to each other by *strings* (completely flexible molecular linkers), or attached to each other by *rods* (completely rigid molecular linkers). There are, in reality, *degrees* of flexibility, and this model replaces real linking groups with two limiting models for them.

(26) Values of the entropies of condensation are tabulated for a wide range of organic colvents in the *CRC Handbook for Chemistry and Physics*.

(27) Simanek, E. E.; Mammen, M.; Gordon, D. M.; Chin, D.; Mathias, J. P.; Seto, C. T.; Whitesides, G. M. *Tetrahedron* **1995**, *51*, 607–619.

(28) Mammen, M.; Simanek, E. E.; Whitesides, G. M. *J. Am. Chem. Soc.* **1996**, *118*, 12614–12623.

Intramolecular motions are of two sorts: those that occur at high frequency and can be modeled accurately as harmonic oscillations and those that occur at low frequency and cannot. First, high-frequency, harmonic motions occur with well-defined maximum and minimum spatial distortions about a single equilibrium position and are characterized accurately by spring constants. Scheraga has previously described these types of motions as 'hard' vibrations.[29] These true vibrational motions include bond stretches, wags, bends, and other motions that correspond to the peaks of a typical infrared spectrum (100−4000 cm^{-1}). Second, most low frequency (<100 cm^{-1}) vibrations are modeled poorly as harmonic oscillations either because the atoms move in highly asymmetric energy wells or move frequently from well to well (and "escape" the restoring force that is characteristic of a vibration). Scheraga has previously called such low-frequency motions "soft" vibrations, and the major contributions to this class of motion are the torsions about single bonds.[29] We define the entropy arising from these soft vibrations as conformational entropy, ΔS_{conf}.

In this simple model, we assume that all accessible conformational space has the same potential energy; that is, the potential energy E (J/mol) is completely independent of the conformation (defined by the coordinates r, θ, and ϕ) of the molecule. Since potential energy can be defined with respect to the energy of the global minimum, it follows that E = constant = 0. The classical (nonquantum mechanical) definition of the partition function for conformational entropy, q_{conf} (eq 7), for a ball then simplifies and is the conformational volume, V_{conf} (Å^3), available to the ball (relative to its own volume, V_{ball} (Å^3)) as it moves while all other balls are fixed in space.[30] The partition function is unitless. The form of eq 7 is

$$q_{conf} = \frac{q_{free}}{q_{fixed}} = \frac{\int\int\int e^{-E(r,\theta,\phi)/kT} dr\, d\theta\, d\phi}{V_{ball}} = \frac{\int\int\int dr\, d\theta\, d\phi}{V_{ball}} = \frac{V_{conf}}{V_{ball}} \quad (7)$$

consistent with the notion that $q_{conf} = 1$ for a ball that is unable to move (i.e., it is rigidly connected to the rest of the molecule). This model is a simplification: it permits us, however, to estimate the *maximum* impact of conformational entropy on self-assembly. The conformational entropy, S_{conf} (J mol^{-1} K^{-1}) is derived from its partition function and is given by eq 8.

$$S_{conf} = R \ln\left(\frac{V_{conf}}{V_{ball}}\right) \quad (8)$$

In this next section, we illustrate the use of eq 8 using an example based on the 4-particle flexible assembly flex(M)$_3$ (**2**) and the 4-particle rigid assembly hub(M)$_3$ (**3**) (Figure 6). The three M components of flex(M)$_3$ (represented by white balls with volume V_M) are connected by three strings of length r_{flex} to a central point. The conformational volume accessible to one ball is approximated by the volume of a sphere of radius r_{flex}: $V_{flex} = (4/3)\pi r_{flex}^3$. Since all three balls are equivalent, the total conformational entropy is given by $S_{conf,flex} = 3R \ln(V_{flex}/V_M)$ (eq 8). The conformational volume accessible to one unit of M in rigid hub(M)$_3$ is given by the volume of the ball itself, V_M: that is, $V_{hub} = V_M$ and $S_{conf} = 3R \ln(V_M/V_M) = 0$. That is, by definition, the conformational entropy of the completely rigid hub(M)$_3$ is zero.

Upon complexation with three components of CA (represented by black balls), we have two choices on how to model the conformational entropy of flex(M)$_3$ *in the complex*.

Method I. If the length of the strings is just sufficient to bridge all the M and CA components together, then there is (in one limit) *no* extra conformational volume accessible to the balls of **2** other than the volume of the balls themselves. In such a case, $S_{conf,2} = 3R \ln(V_M/V_M) = 0$. The total change in conformational entropy for the assembly of flex(M)$_3$ and 3CA is then $\Delta S^{assoc}_{conf,2} = -3R \ln(V_{flex}/V_M)$. From realistic values of V_M (80 Å^3) and V_{flex} (6000 Å^3), we calculate that the value of $\Delta S^{assoc}_{conf,flex} = -110$ J mol^{-1} K^{-1} (∼32 kJ/mol at 25 °C).[31]

Method II. If the length of the strings is much longer than is required to bridge all the M and CA components together, then there is conformational volume available to the balls of **2**. This extra volume accounts for any *residual entropy* due to motions in the aggregate. The total accessible volume in the complexed flex(M)$_3$ can be crudely approximated as the same as for one M in the uncomplexed flex(M)$_3$. That is, we assume that the motions of the three M components in uncomplexed flex(M)$_3$ occur independently, but the motion in the aggregate is coupled. The volume of the complex of 3 balls that move as a single unit is $3V_M$. The total conformational entropy in the complexed flex(M)$_3$:3CA is then $\Delta S^{assoc}_{conf,flex:3CA} = -3R \ln(V_{flex}/V_M) + R \ln(V_{flex}/3V_M)$. From realistic values of V_M (80 Å^3) and V_{flex} (6000 Å^3), we calculate that the value of $\Delta S^{assoc}_{conf,flex} = -80$ J mol^{-1} K^{-1} (∼24 kJ/mol at 25 °C). We believe that this second method more closely represents reality than the first method. From NMR titration and competition experiments reported previously, we estimate that the difference in ΔG of assembly between **2** and **3** is greater than 10−15 kJ/mol, which is consistent with the theoretical estimate here.[27]

Figure 6 illustrates the calculations of all entropic contributions to the assembly of **2** and **3**. In addition to the values of conformational entropies just estimated, we use the expressions derived earlier for the corrected translational entropy (eq 6b) and the rotational entropy (eq 3c). In summary, the difference in the value of ΔS^{assoc}_{total} between **2** and **3** is due to differences in ΔS^{assoc}_{conf} and equals ∼24 kJ/mol at 25 °C (∼6 kcal/mol). This estimate further accounts for the significant residual

(29) Go, N.; Scheraga, H. A. *Macromolecules* **1981**, *9*, 535.

(30) Formally, the partition function should be an integral over momenta as well potential energy. Since the temperature in our system is assumed to be constant, the kinetic energy component (momenta) of the partition function will be constant and is therefore ignored in our analysis.

(31) As a very rough check on this estimate, we employ a simple rule of thumb regarding the restriction of torsional motion around a single bond. If a single bond has three equal torsional energetic minima available to it prior to complexation, and one torsional minimum afterward, then the cost of restricting that torsional motion is approximately $R \ln(3/1) = 9$ J mol^{-1} K^{-1}. There are 8 unrestricted single bonds on each of the arms of hub(M)$_3$ that become (in one extreme case) restricted on complexation with 3CA. The total change in conformational entropy on complexation can be crudely estimated by the *total* number of bonds that become restricted. There are 3 equal arms on hub(M)$_3$, with 8 torsions per arm, giving 24 torsions. The total entropic cost of complexation is therefore estimated using this crude treatment to be $24 \times R \ln(3) = 220$ J mol^{-1} K^{-1}. This value is twice that predicted by the B−R&S model.

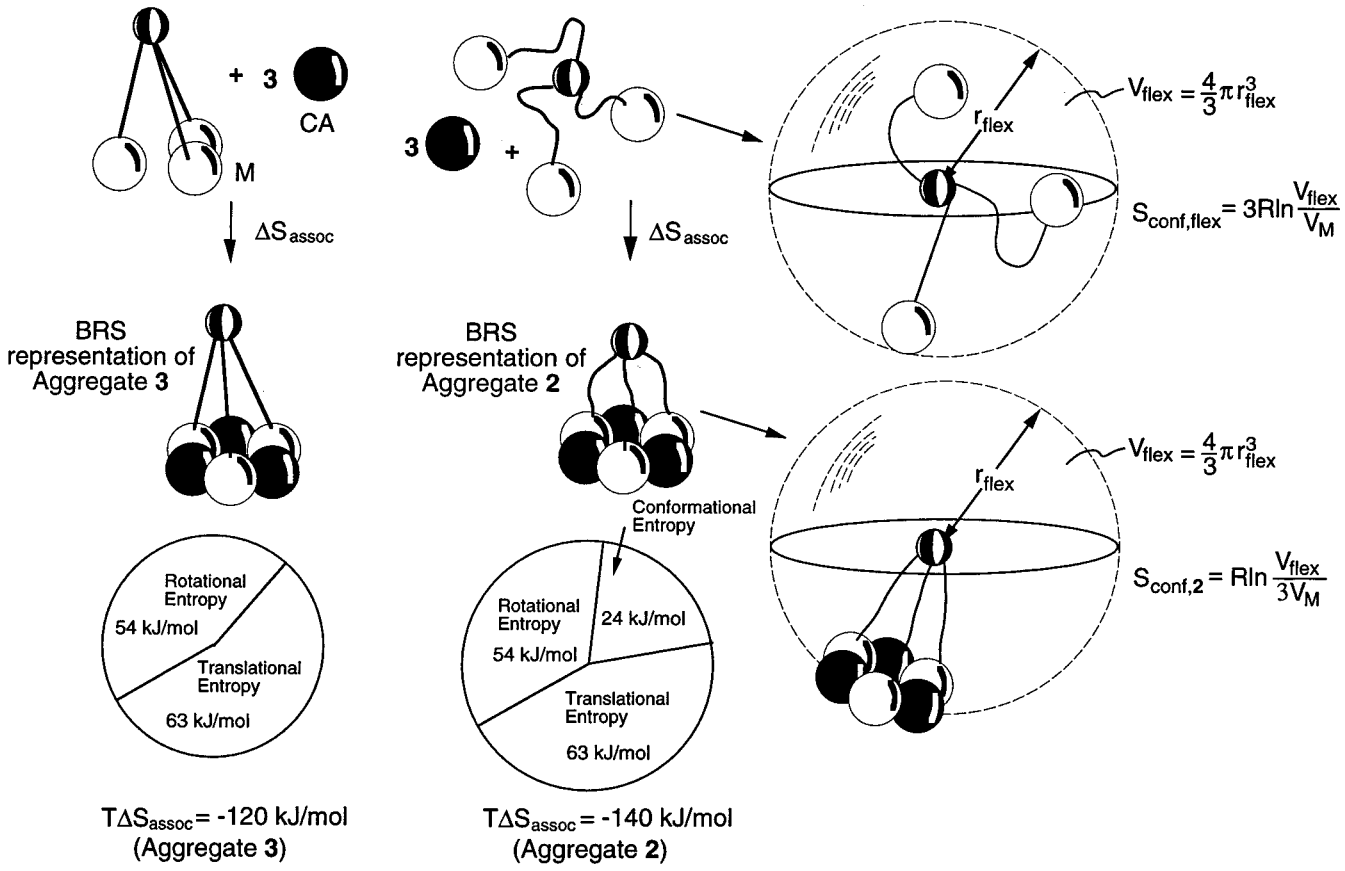

Figure 6. Entropy of association of the flexible 4-particle aggregate **2** and the rigid 4-particle aggregate **3**. These two aggregates and their components are approximated as balls attached by rods and strings. The changes in entropy upon assembly are calculated using eqs 1, 3c, 6b, and 8.

entropy (~8 kJ/mol to ~2 kcal/mol) due to internal motions in the formed aggregate **2**; that is, the (BRS) model does *not* have to assume that all intramolecular motions stop on complexation.

Conclusions

The existing theoretical models for estimating enthalpies of assembly for noncovalent aggregates are growing in accuracy.[6] Furthermore, the collective intuition of chemists about enthalpy is strong. Entropy is as important in noncovalent assembly as enthalpy; models for predicting its various components are highly inaccurate or do not exist. Furthermore, intuition is weak regarding when and to what degree changes in entropy are important. *This work outlines a useful, if semiquantitative, model for estimating the changes in all components of entropy during multiparticle assembly.* The new contributions of this work are methods for estimating translational and conformational entropy of multiparticle assembly; we believe that these methods will be valuable to the community involved with molecular recognition.

The correction to the Sackur−Tetrode equation to estimate translational entropy in solution using free volume theory is an extension of the previously described model for free volume. The free volume model is not, however, broadly employed. Both it and its extension described here may be useful to those estimating translational entropy in a condensed phase. This correction is approximate, but useful and easy to use. This model provides a significantly improved and physically more reasonable estimate of translational entropy in solution than does the Sackur−Tetrode equation, which has been used in uncorrected form in all estimations of molecular recognition in solution of which we are aware.[7,8] Corrections to the Sackur−Tetrode equation based on free volumes of irregularly shaped molecules may represent second-order solutions to the problem of estimating translational entropy in solution.

The BRS model for estimating changes in conformational entropy is also approximate and is a limiting model; that is, it generally overpredicts the difference in entropy between two aggregates. The BRS model is, however, readily applied (it is based on determining volumes of simple geometric shapes) and might find use mainly due to the lack of existing models for this important component of the total entropic cost of multiparticle assembly.

The models presented here may be collectively useful in designing molecules that bind polyvalently to polyvalent targets—for example, pharmaceuticals presenting multiple ligands and intended to bind to cellular surfaces containing multiple receptors.[32−35] The strategy of polyvalency works only if the total entropy of binding a polyvalent molecule containing *N* ligand groups to aggregate of *N* receptors is greater than the entropy of

(32) Haywood, A. M. *J. Virol.* **1994**, *68*, 1−5.
(33) Lentz, T. L. *J. Gen. Virol.* **1990**, *71*, 751−766.
(34) Albelda, S. M.; Smith, C. W.; Ward, P. A. *Faseb J.* **1994**, *8*, 504−512.
(35) Mammen, M.; Dahmann, G.; Whitesides, G. M. *J. Med. Chem.* **1995**, *38*, 4179−4190.

binding *N* individual ligands to the same aggregate of *N* receptors. Physical tethering of the ligands reduces the translational and rotational entropies of binding but, depending on the nature of the tethers (rigid or flexible), introduces conformational entropy. Whether a polyvalent design is successful or not depends on the delicate balance between the conformational entropy added and the sum of translational and rotational entropies removed. This work provides semiquantitative and conceptual models for both sides of this balance. Incomplete understanding of entropy in the design of polyvalent inhibitors has resulted in many synthetic polyvalent molecules that are only marginally more effective than their monovalent counterparts. Bivalent systems joined by flexible linkers (e.g. oligoethyleneglycol or polymethylene)[36] provide examples of systems that can almost be guaranteed to fail for entropic reasons.

(36) Glick, G. D.; Knowles, J. R. *J. Am. Chem. Soc.* **1991**, *113*, 4701−3.

Some of the ideas outlined here have been described earlier in pieces, often separated by decades in the literature; others (including the extension of translational entropy to solutions and the BRS model for conformational entropy) are new. We have applied the model to a system—structures based on CA·M—that we are familiar with, and the estimates are consistent with experiment.[28] The model should prove useful both numerically and conceptually to the community interested in molecular recognition.

Acknowledgment. This work was supported by NIH Grants GM 30367 and GM 52126NSF and NSF Grant CHE 91-22331. The NMR facilities at Harvard were supported by NIH Grant 1-S10-RR04870-01 and NSF Grant CHE88-14019. M.M. gratefully acknowledges Eli Lilly for a predoctoral fellowship.

JO970944F