CHEMICAL REVIEWS

Volume 96, Number 8 December 1996

Comparisons of Indefinite Self-Association Models

R. Bruce Martin

Chemistry Department, University of Virginia, Charlottesville, Virginia 22903

Received March 19, 1996 (Revised Manuscript Received August 14, 1996)

Contents

I. Introduction

Stacking interactions contribute importantly to the stability of secondary structure in nucleic acid polymers such as DNA. Stacking interactions among planar aromatic molecules such as purines and pyrimidines are now generally accepted as a significant component of their interactions. Stacking and other indefinite self-association reactions occur in many types of molecules including drugs and dyes. This article develops general equations for two models that deal with molecules that undergo self-association into structures of indefinite size. The general equations are then applied to several experimental techniques. The development is of wide general applicabilty.

Early studies employed osmometry or sedimentation equilibria to evaluate the extent of self-association. These studies, limited in number, were followed by many using chemical shifts from nuclear magnetic resonance spectroscopy. These three techniques provide complementary information: osmometry, number average properties; sedimentation, weight average molecular weight; and NMR chemical shifts, direct concentration information without concern for nonideality. In this article results from these three methods and others will be compared.

Results have usually been interpreted in terms of the isodesmic model of indefinite stacking wherein the addition of a molecule to a stack occurs with the identical free energy and equilibrium constant, *K*, as for addition of previous molecules. This model is termed the *equal K (EK)* model. Alternatively, it has been suggested that though the enthalpy of addition to a growing stack may be constant, successive

S0009-2665(96)00037-4 CCC: $25.00

Bruce Martin is Professor Emeritus of Chemistry at the University of Virginia, where he joined the faculty in 1959. He was born in Chicago and earned his B.S. degree in chemistry at Northwestern University in 1950 and his Ph.D. degree in photochemistry at the University of Rochester in 1953. He held postdoctoral appointments at both Caltech and Harvard, and has twice been on leave from Virginia at Oxford University. Dr. Martin has published more than 200 articles in the fields of biophysical and bioinorganic chemistry.

additions should be increasingly less probable and hence less favored entropically, and successive equilibrium constants should taper off in value. This model is termed the *attenuated K (AK)* model.[1] This model is seldom employed and for the first time this article applies it to NMR data. Both models may be refined by allowing the equilibrium constant for dimer formation to differ from the value of the equilibrium constant for subsequent additions. Herein, for the first time, both models, with distinguished dimer equilibrium constants, are applied to all three experimental methods, and with the inclusion of nonideality.

The designations unstacked and stacked both describe an ensemble average over many conformations all in rapid equilibrium. The designations unstacked and stacked should not be assigned to specific conformations, although there are likely fewer stacked than unstacked conformations. It is possible that for the same compound a difference in equilibrium constants between two experimental methods is real: the methods may sense differently the ensemble of conformations. For example, osmometry, a number average method that counts moles, may sense association among molecules that are not registered by upfield shifts in NMR spectra, which require a more specific geometry and short-range interactions. Thus the equilibrium constant for stacking as calculated by osmometry would be greater than that calculated by NMR.

This article deals only with equilibrium or thermodynamics and not with mechanism or kinetics. Although addition to a growing stack is framed as sequential in both the EK and AK models, stack formation may well occur by random association. For example, in addition to adding monomer to pentamer, a hexamer may form by stacking of two trimers or by stacking of a dimer and tetramer. Observation in kinetics experiments of a single relaxation time without broadening suggests that random association dominates over the sequential pathway.[2] The lifetime of a stacked complex is only about 20 ns. Since we are concerned only with equilibrium and write expressions for formation of all complexes, their favored route of formation is not a factor in our analyses.

In the usual treatment from NMR spectra, only nearest-neighbor interactions are considered and the chemical shift of a molecule at the end of a stack is taken as half the upfield chemical shift of monomer to a molecule within a stack. However, we relax that requirement and let the ratio vary. In an entirely separate and new analysis we allow for a contribution to the chemical shift of next to nearest-neighbor interactions. Both NMR treatments are conducted with both the EK and AK models.

Several aspects of the models employed in this article require evaluation of more parameters than usual. Precise data is needed. For objective analysis the data have been subjected to a nonlinear least-squares treatment. The analysis is performed on equations written with the most error-prone observable alone on the left hand side and all other variables, constants, and parameters to be fitted on the right. The best fit minimizes the variance in the most error-prone observable. The analysis assumes no error in the concentration variable. While the most error-prone observable is usually some sort of meter reading, results are not reported so primitively. In all three types of experiments mainly considered here, reported results are proportional or nearly so to the most error-prone observable in the experiment. Thus the left hand side of the equation for sedimentation is set up with the weight average molecular weight, for vapor pressure osmometry with the osmotic coefficient, and for NMR with chemical shifts. A reasonable weighting of points is a major advantage of nonlinear least-squares of the most error-prone observable over casting equations into a form appropriate for linear plots. In the latter procedure linearizing may weight points at one end of a plot much more than those at the other, sometimes hundreds of times more. In only one case analyzed here were the error limits on individual points specified; in the others the errors were taken as absolute rather than fractional, in general agreement with error bars and remarks in some papers. A widely available computer program has been used for the nonlinear least-squares analysis.[3] Of course, a good nonlinear least-squares fit does not guarantee that the model is correct; with enough adjustable parameters almost any model will fit. The two models given below usually yield closely similar goodness of fits, suggesting that the fits are not constrained by too limited a number of parameters and are down to random errors in the experimental observables. Discoveries have been made by application of nonlinear least-squares: cooperative proton as well as calcium binding was found in CaATPase.[4]

II. General Indefinite Self-Association Formulation

The indefinite self-association of solute A may be described by the following successive equilibria and associated equilibrium constants with the brackets representing molar concentrations. (The setup with molality would be similar.)

$$A + A \rightleftharpoons A_2 \qquad K_2 = [A_2]/[A]^2 \tag{1}$$

$$A_2 + A \rightleftharpoons A_3 \qquad K_3 = [A_3]/[A_2][A] = [A_3]/K_2[A]^3$$

$$A_3 + A \rightleftharpoons A_4 \qquad K_4 = [A_4]/[A_3][A] = [A_4]/K_2K_3[A]^4$$

$$A_{i-1} + A \rightleftharpoons A_i$$

$$K_i = [A_i]/[A_{i-1}][A] = [A_i]/(K_2...K_{i-1})\,[A]^i$$

where $i = 2$ to infinity. The total molar concentration of component A in all its forms is given by

$$C_T = [A] + 2[A_2] + 3[A_3] + 4[A_4] + ...$$
$$= [A]\,(1 + 2\,K_2[A] + 3\,K_2K_3[A]^2 + 4\,K_2K_3K_4[A]^3 + ...) \tag{2}$$

We define the mole fraction of monomer as

$$\alpha = [A]/C_T = 1/(1 + 2K_2[A] + 3K_2K_3[A]^2 + 4K_2K_3K_4[A]^3 + ...) \tag{3}$$

The weight fraction of higher n-mers is given by

$$\alpha_2 = 2[A_2]/C_T \quad \alpha_3 = 3[A_3]/C_T \quad \alpha_4 = 4[A_4]/C_T \quad \alpha_i = i[A_i]/C_T \tag{4}$$

all of which may be expressed in terms of the mole fraction monomer and equilibrium constants.

If instead of an indefinite association one has dimer, trimer, or a cap at any n-mer, the equations developed here may not be used. In the terminology used below, large values for ρ or τ do not result in reduction of the equations to those of a dimer. Capped n-mers require their own analysis. Since the most common specific n-mer is the dimer, we introduce equations for the dimer in each situation. The general dimer equation results from combining eq 1, the first two terms on the right of eqs 2 and 3, and $L = K_2C_T$ to obtain

$$\alpha(1 + 2\alpha L) = 1 \tag{5}$$

from which the mole fraction of monomer in a solution containing only dimer and monomer is

$$\alpha = [-1 + \sqrt{(8L + 1)}]/4L \tag{6}$$

At the point where $L = 1$ we find $\alpha = 0.50$. Both eqs 5 and 6 are useful in simplifying dimer expressions below. Solving eq 5 for L we also find

$$L = (1 - \alpha)/2\alpha^2$$

At this juncture the analysis of indefinite self-association diverges, depending upon the choice of model.

A. Equal *K* (EK) Model

In the simplest and most common usuage the equal model asssigns all equilibrium constants for self-association as equal:

$$K_E = K_2 = K_3 = K_4 = ... = K_i \quad \text{with } i = 2 \text{ to infinity} \tag{7}$$

With this equality and setting two dimensionless variables

$$x = K_E[A] \quad \text{and} \quad L = K_EC_T \tag{8}$$

eq 2 becomes

$$L = x(1 + 2x + 3x^2 + 4x^3 + ...) = x/(1 - x)^2 \tag{9}$$

the last equality resulting from the series expansion

$$(1 - x)^{-2} = 1 + 2x + 3x^2 + 4x^3 + ...\,(i + 1)x^i \quad \text{for } x < 1$$

In these systems we always find $x < 1$.

Equation 9 possesses the great merit of capturing an infinite series in terms of a simple quotient. Equation 9 is quadratric in x and solution yields

$$x = (2L + 1 - \sqrt{4L + 1})/2L$$

with x ranging from near zero at low L, through 0.25 at $L = 2$, to near but always less than unity at high L.

The mole fraction of monomer follows from eq 3 as

$$\alpha = [A]/C_T = x/L = (1 - x)^2 = (1 - \alpha L)^2 \tag{10}$$

from which

$$\alpha = \frac{2L + 1 - \sqrt{4L + 1}}{2L^2} \tag{11}$$

Equation 11 is an important equation from which α may be determined at a given K_E and molar concentration as $L = K_EC_T$, when $\rho = 1$ of the next paragraph. Solving eq 10 for L we also find $L = (1 - \sqrt{\alpha}/\alpha = \alpha L/(1 - \alpha L)^2$.

We now allow for the possibility that the equilibrium constant for the first step in a self-association—dimer formation—differs from the remaining constants.[5] We define the ratio $\rho = K_2/K_E$ so that eq 7 becomes

$$K_E = K_2/\rho = K_3 = K_4 = ... = K_i \tag{12}$$

If the dimer forms more easily than subsequent additions, $\rho > 1$, while if the dimer forms with more difficulty, $\rho < 1$, and if there is a nucleation process involved, $\rho \ll 1$. Steric hindrance and electrostatic repulsion in charged molecules may lead to $\rho > 1$. Since a dimer is made from two monomers that lose some orientational freedom while addition to a stack involves loss of entropy from only one monomer, we might expect $\rho < 1$ for neutral, sterically unhindered monomers.[6] Substitution in eq 2 now leads in place of eq 9 to

$$L = x(1 + \rho\,(2x + 3x^2 + 4x^3 + ...)) = x(1 - \rho) + \rho x/(1 - x)^2 \tag{13}$$

The last equation is one of several forms that reduce to eq 9 when $\rho = 1$. Equation 13 is cubic in x, and

finding the value of x involves solving a cubic equation either analytically or by iterations on a computer.

The mole fraction of monomer now becomes

$$\alpha = (1 - x)^2/(1 - x(2 - x)(1 - \rho))$$

which reduces to eq 10 when $\rho = 1$. Noting that with $x = \alpha L$ and rearranging we obtain in place of the quadratic eq 10

$$\alpha^3 L^2(\rho - 1) + \alpha^2 L(L - 2(\rho - 1)) - \alpha(2L + 1) + 1 = 0 \quad (14)$$

This cubic equation is routinely solved for α in this research. Surprisingly, sometimes two roots occur in the limited range of $0 < \alpha < 1$; exhaustive analysis reveals that the lesser of the two roots is always the desired one.

Once the values of ρ and K_E have been determined, the weight fraction of higher n-mers as defined in eq 4 is given in the EK model by

$$\alpha_i = \rho i[A]^i {K_E}^{i-1}/C_T = \rho i \alpha x^{i-1} = \rho i \alpha^i L^{i-1} \quad \text{with } i \geq 2 \quad (15)$$

B. Attenuated *K* (AK) Model

In the attenuated model successive equilibrium constants for self-association taper off according to the prescription $K_{Ai} = K_A/i$ to give

$$K_2 = K_A/2,\ K_3 = K_A/3,\ K_4 = K_A/4, \text{ etc.} \quad (16)$$

or $K_A = 2K_2 = 3K_3 = 4K_4 = \ldots = iK_i$.[1] As in eq 8 we introduce the convenient dimensionless variables

$$x = K_A[A] \quad \text{and} \quad L = K_A C_T \quad (17)$$

and from substitution in eq 2 deduce that

$$L = x\,(1 + x + x^2/2! + x^3/3! + \ldots + x^i/i!) = xe^x \quad (18)$$

Once again an infinite series is contained in a simple expression which, however, requires an iterative procedure for its solution. Since $\alpha = x/L$, we also find for the mole fraction of monomer, $\alpha = e^{-x} = e^{-\alpha L}$, when $\tau = 1$ of the next paragraph.

We may allow for the equilibrium constant for dimer formation not following the pattern of eq 16 and, analogous to ρ in the previous section, define the ratio $\tau = 2K_2/K_A$ so that

$$K_2 = \tau K_A/2,\ K_3 = K_A/3,\ K_4 = K_A/4, \text{ etc.} \quad (19)$$

or $K_A = 2K_2/\tau = 3K_3 = 4K_4 = \ldots = iK_i$. A value of $\tau > {}^2/_3$ favors dimer formation. Substitution into eq 2 leads to

$$L = x(1 + \tau(x + x^2/2! + x^3/3! + \ldots + x^i/i!)) = x\,(1 + \tau(e^x - 1)) \quad (20)$$

When $\tau = 1$, eq 20 reduces to eq 18.

From eq 20 we find for the mole fraction of monomer

$$\alpha = x/L = 1/(1 + \tau(e^x - 1)) \quad (21)$$

which is also solved by iteration.

We apply eq 20 to eq 4 to find for the mole fraction of the ith species

$$\alpha_i = \tau\alpha x^{i-1}/(i-1)! = \tau\alpha^i L^{i-1}/(i-1)! \quad \text{with } i \geq 2 \quad (22)$$

C. Other Models

Several other models have been proposed[7] and, although less useful than the EK and AK models, we cast them in the more convenient format of this article.

In the increasing equilibrium constant model the successive constants increase according to the prescription $K_i = K(i - 1)/i$, so that $K_2 = K/2$, $K_3 = 2K/3$, $K_4 = 3K/4$, etc., or $K = 2K_2 = 3K_3/2 = 4K_4/3$, etc. Substitution into eq 2 yields with the terminology already developed

$$L = KC_T = x(1 + x + x^2 + x^3 + \ldots) = x/(1 - x)$$

from which $\alpha = 1 - \alpha L$ and $1/\alpha = L + 1$. We now introduce a variable $K_2 = \theta K/2$, with θ an additional variable, so that the more general expression for L becomes

$$L = x(1 + \theta(x + x^2 + x^3 + \ldots)) = x(1 - x(1 - \theta))/(1 - x)$$

The equation for α now becomes quadratic. Mathematically this is the simplest model but, even after inclusion of a variable K_2, its use is limited to systems with a mild cooperativity.

In what might be called the middling K model the successive constants mildly decrease according to $K_i = K_M(i - 1)/(i - 2)$ for $i \geq 3$, so that $K_3 = 2K_M$, $K_4 = 3K_M/2$, $K_5 = 4K_M/3$, etc., or $K_M = K_3/2 = 2K_4/3 = 3K_5/4$, etc. We also define $K_2 = \varphi K_M/2$ with φ as an additional variable. If $K_2 = K_3$ then we have $\varphi = 4$. Substitution of these quantities into eq 2 yields

$$L = K_M C_T = x(1 + \varphi x(1 + 3x + 6x^2 + 10x^3 + \ldots)) = x(1 + \varphi x/(1 - x)^3)$$

The corresponding expression for α is a quartic equation.

Both the middling K model and AK model begin similarly with $K_3/K_4 = {}^4/_3 = 1.33$. Subsequently the equilibrium constants fall off more rapidly in the AK model, becoming in the limit of long chains $K_\infty = 0$ in the AK model and $K_\infty = K_M$ in the middling K model. Since the middling K model yields a quartic equation and since the limiting equilibrium constant for infinite chain length is finite and not zero, we do not consider the model further. The model has been applied to the self-association of methylene blue.[8]

III. Nonideality

For neutral molecules the activity coefficient that allows for nonideality begins with a value of unity

in the most dilute solutions and increases with concentration and molecular size. For associating molecules it is highly convenient to write the activity coefficient γ in the form

$$\ln \gamma_i = BM_iC_T = iBM_1C_T \tag{23}$$

where γ_i is the activity coefficient of the *i*-mer, the molecular weight of an associated *i*-mer is given by $M_i = iM_1$, where M_1 is the molecular weight of the monomer, and C_T is the total molar concentration.[9,10] Equation 23 states that upon association the contribution to the activity coefficient due to excluded volume remains constant on a per weight basis. For nonelectrolytes ln γ_i is positive and the second virial coefficient $B > 1$. Association reactions reduce the apparent B, yielding negative values in some cases.

The form of eq 23 is especially advantageous when applied to equilibria because the activity coefficients cancel in the activity formulation of equilibrium constants so the values are numerically the same as the concentration constants in the bank of equations after eq 1. For example, for the dimerization reaction with a_D and a_M the activity of dimer and monomer, respectively, we have

$$K_2 = \frac{a_D}{a_M{}^2} = \frac{[A_2]\gamma_2}{[A]^2\gamma_1{}^2} = \frac{[A_2]e^{2BM_1C}}{[A]^2(e^{BM_1C})^2} = \frac{[A_2]}{[A]^2}$$

Even though γ cancels in the numerator and the denominator of the equilibrium constants, it still appears in other ways in the results of both sedimentation and osmotic coefficient experiments.

There has been a lack of consistency in applying units to the terms in eq 23. Since the left hand side is dimensionless, so must be the right. With the molecular weight in grams/mole, and the concentration C_T in moles/liter, the units of *B* become liters/gram. When the concentration is expressed in grams/liter, this corresponds to the product M_1C_T. It is awkward to list the product BM_1 as liters/gram, as the units of *B* become mole (liter/gram^2), the meaning of which is difficult to fathom. The units of the product BM_1 are liters/mole, and this product increases as the monomer molecular weight increases.

The units of the second virial coefficient *B* are appropriately volume per unit mass, preserving its value as a proportionality constant to first order, independent of size and concentration, in eq 23. Since the density of organic molecules is of the order of 1 g/mL, we expect the value of *B* to be of the order of 1 mL/g, or 10^{-3} L/g. Since the milliliter/gram scale provides more conveniently sized quantities, we quote *B* values in these units. Assuming no association, from osmotic coefficient data for sucrose[11] we deduce $B = 0.55$ mL/g.

As shown in the next section, it is crucial to consider the nonideal term in sedimentation experiments as it alone accounts for maxima in plots of apparent weight average molecular weight versus concentration. The appearance of a nonideal term in osmotic coefficient results is less dramatic, and its presence has almost always been ignored. However, a proper formulation requires that it be included. Since the osmotic coefficient results are often over a limited concentration range or lack the requisite precision, nonlinear least-squares refinement of the results often fails to produce reliable values of B. We use a prototypical value of 0.70 mL/g in such cases. Since the NMR chemical shifts are measures of weighted average concentrations, the nonideal term does not enter into the NMR results.

IV. Sedimentation Equilibrium

The usual equation for evaluating equilibrium sedimentation results is

$$\frac{M_1}{M_{wa}} = \frac{M_1}{M_{wc}} + BM_1C_T \tag{24}$$

where M_{wa} is the apparent weight average molecular weight, M_1 is the monomer molecular weight, M_{wc} is the weight average molecular weight allowing for asssociation but not nonideality, and BM_1C_T is the nonideal term discussed in section III. The ratio that incorporates association is given by[10]

$$\frac{M_{wc}}{M_1} = \frac{[A]\, dC_T}{C_T\, d[A]} = \frac{x\, dC_T}{C_T\, dx} \tag{25}$$

where *x* is as defined in eqs 8 and 17, depending on the model.

Since we wish to minimize the variance in the most error-prone observable, the apparent weight average molecular weight, we rearrange eq 24 to obtain

$$M_{wa} = \frac{M_1}{(M_1/M_{wc}) + BM_1C_T} \tag{26}$$

where the first term in the denominator, always <1, depends upon the model and is obtained as a reciprocal from the differentiation indicated in eq 25. The second term in the denominator increasingly offsets the first as the concentration increases.

Application of eq 25 to a dimer only yields

$$M_{wc}/M_1 = \alpha(1 + 4\alpha L) = 2 - \alpha$$

the last equality resulting from application of eq 5.

A. EK Model

To substitute for the first term in the denominator of eq 26 we apply the prescription of eq 25 to eq 9 with $\rho = 1$ to obtain $M_{wc}/M_1 = (1 + x)/(1 - x)$. For the more general case where the value of the first equilbrium constant may differ from the value for the subsequent constants, $\rho \neq 1$, we apply eq 25 to eq 13 to find

$$\frac{M_{wc}}{M_1} = \frac{(1-x)^2(1-\rho) + \rho(1+x)/(1-x)}{1 - x(2-x)(1-\rho)} = \alpha((1-\rho) + \rho(1+x)/(1-x)^3) \tag{27}$$

Equations 14, 26, and 27 are used in the nonlinear least-squares analysis.

B. AK Model

Similar to the treatment above we apply eq 25 to eq 18 with $\tau = 1$ to obtain $M_{wc}/M_1 = 1 + x$. For the

Table 1. Sedimentation Equilibrium Stacking Results

			equal K model (EK)				attenuated K model (AK)			
solute	points[a]	range[b]	K_E	ρ	B	σ^c	K_A	τ	B	σ^c
purine[d]	57	1.0 M	2.76(2)	0.94(2)	1.20(2)	1.9	10.4(1)	0.38(1)	0.79(2)	1.7
inosine[e]	12	0.11 m	2.6(3)	0.8(1)	0.7 set	4.0	9(1)	0.42(9)	0.7 set	4.1
deoxyadenosine[e]	19	0.05 m	10.7(6)	0.9(2)	0.7 set	9.6	41(3)	0.39(9)	0.7 set	9.8
5′-H(AMP)$^{-}$[f]	54	0.60 M	5.31(4)	1.92(4)	0.82(1)	3.7	18.3(2)	0.99(4)	0.63(1)	4.9
cytidine[g]	32	0.66 M	1.13(4)	0.67(3)	0.70(5)	1.5	3.5(2)	0.40(3)	0.44(5)	1.5

[a] Number of experimental points. [b] Upper limit of reported concentration range in molarity (M) or molality (m). [c] Standard deviation from nonlinear least-squares fit in molecular weight units. Numbers in parentheses for three determined parameters indicate one standard deviation in last digit(s). [d] From ref 13. [e] From ref 20. [f] Monoprotonated, negatively charged phosphate form from 5.0 < pH < 5.3 from reference 14. [g] From ref 15.

more general case with $\tau \neq 1$ we apply eq 25 to eq 20 to find with eq 21

$$\frac{M_{wc}}{M_1} = \frac{1 + \tau(e^x(1 + x) - 1)}{1 + \tau(e^x - 1)} = \alpha(1 + \tau(e^x(1 + x) - 1)) \quad (28)$$

Equivalent equations have been published in an alternative formulation.[12] Equations 21, 26, and 28 are used in the nonlinear least-squares analysis.

C. Results

Plots of M_{wa}/M_1 versus molar concentration for purine[13] and 5′-H(AMP)$^{-}$ [14] exhibit maxima at ordinate values greater than 2. These plots emphasize two features: n-mers higher than dimers are formed, and the nonideal B term in the denominator of eq 26 becomes important at the higher concentrations. The leveling off at the highest concentrations is accounted for entirely by the nonideal term BM_1C_T in the denominator of eq 26. With such pronounced curvature the B term is relatively well determined in sedimentation experiments.

Table 1 presents nonlinear least-squares analysis for three parameters each in the EK and AK models evaluated directly from the reported sedimentation weight average molecular weights. Since $\rho = 0.94$-(2) or nearly unity for purine, there is good agreement for K_E and B with the careful original study that in effect set $\rho = 1.00$.[13] Since for cytidine $\rho = 0.67(3)$, there is lesser agreement with the original paper that reported $K_E = 0.82$ M^{-1} and $B = 0.35$ mL/gram.[15] From the point of view of goodness of fit there is no clear cut preference between the models; except for 5′-H(AMP)$^{-}$, with a monoprotonated, negatively charged phosphate, the standard deviation from the nonlinear least-squares fit is closely similar in the two models. Curiously, the AK model gives the poorer fit despite the fact that its ever decreasing equilibrium constants would seem better able to provide for additions of a negatively charged molecule to a negatively charged stack. For ionic species, sedimentation equations are found to be of the same form as eq 24.[16] The relatively high $\rho > 1.0$ and $\tau > 0.67$ values for 5′-H(AMP)$^{-}$ suggest that dimer formation is relatively more favored than subsequent additions to the stack, suggestive of steric hindrance, but more probably due to electrostatic repulsion in this case. However, some cautions are in order because of the negative charge on the phosphate of 5′-H(AMP)$^{-}$, even if remote from the stacking base. It has been suggested that eq 23 may be applied to charged molecules,[9] and this conclusion is supported by the reasonable B values for 5′-H(AMP)$^{-}$ in Table 1. Comparison of these sedimentation results appears in the NMR section IX.D.7.

V. Osmotic Coefficients

In constrast to sedimentation, which most directly evaluates a weight average molecular weight, osmotic measurements are colligative—count moles, and furnish a number average molecular weight. Without including nonideality, the osmotic coefficient, Φ', is the ratio of the number of moles per liter, C_N, to the total molar concentration of a compound in all its forms, C_T, the latter already given by eq 2.

$$\Phi' = \frac{C_N}{C_T} = \frac{[A] + [A_2] + [A_3] + [A_4] + ... + [A_i]}{[A] + 2[A_2] + 3[A_3] + 4[A_4] + ... + i[A_i]} \quad (29)$$

The mean number of monomers per stack is given by $1/\Phi'$, and the number average molecular weight by $M_n = M_1/\Phi'$. Inclusion of nonideality leads to the complete equation for the osmotic coefficient[17]

$$\Phi = \Phi' + BM_1C_T/2 \quad (30)$$

where the nonideality term on the right is discussed in section III. Equation 30 shows that the nonideal term serves to increase the osmotic coefficient while association decreases it. These simple offsets make it possible to obtain good fits without considering nonideality. Without some allowance for nonideality, equilibrium constants for self-association obtained from osmometry will be underestimated.

From the first two terms on the right of eq 29, for the osmotic dimer only we derive

$$\Phi_D' = \alpha(1 + \alpha L) = (1 + \alpha)/2$$

and from eq 6 find

$$\Phi_D' = (4L - 1 + \sqrt{8L + 1})/8L$$

Useful for estimating the dimerization constant at a single concentration is the equation resulting from combining the first two terms of eqs 2 and 29 to yield

$$K_2 = (1 - \Phi')/C_T(2\Phi' - 1)^2$$

Formulation of the association equations for Φ' appropriate to the EK and AK models involves finding expressions for the numerator of eq 29, C_N.

A. EK Model

For the equal K equilibrium model with the substitutions of eqs 7 and 8 we obtain

$$C_N = [A]\,(1 + x + x^2 + x^3 + x^4 + ... + x^i) = [A]/(1 - x)$$

the last equality resulting from the series expansion $(1 - x)^{-1} = 1 + x + x^2 + x^3 + x^4 + ... + x^i$, for $x < 1$. Once again an infinite series gains expression as a simple quotient. Since the mole fraction of monmer $\alpha = [A]/C_T$, and $x = \alpha L$, we may write

$$\Phi' = \frac{C_N}{C_T} = \frac{\alpha}{1 - \alpha L} = \frac{x}{(1 - x)L} \qquad (31)$$

Applying the result of eq 9 we also obtain

$$\Phi' = 1 - x = 1 - \alpha L = \sqrt{\alpha} = \frac{-1 + \sqrt{4L + 1}}{2L}$$

Solving the ends of this equation for $L = K_E C_T$ we obtain

$$K_E = (1 - \Phi')/C_T\Phi'^2$$

useful for estimating K_E at a single concentration and as first derived in an earlier paper.[18]

These equations apply only when $\rho = 1$. For $\rho \neq 1$ we determine

$$C_N = [A](1 + \rho\,(x + x^2 + x^3 + x^4 + ... + x^i)) = [A]\,(1 + \rho x/(1 - x))$$

from which

$$\Phi' = \frac{\alpha(1 + \alpha L(\rho - 1))}{1 - \alpha L} \qquad (32)$$

which reduces to eq 31 when $\rho = 1$. Equations 14, 30, and 32 are used to evaluate by nonlinear least-squares three parameters in the EK model.

B. AK Model

For the attenuated equilibrium model, combination of eqs 16−18 and 29 leads for the number of moles per liter

$$C_N = [A](1 + x/2 + x^2/3! + x^3/4! + ...) = [A](e^x - 1)/x = (e^x - 1)/K_A \quad (33)$$

from which $\Phi' = (e^x - 1)/L = (1/\alpha - 1)/L$.

Equation 33 applies only when $\tau = 1$; for $\tau \neq 1$ we use eq 19 and find

$$C_N = [A](1 + \tau\,(x/2! + x^2/3! + x^3/4! + ...)) = [A](1 + \tau\,(e^x - 1 - x)/x)$$

from which

$$\Phi' = \alpha(1 + \tau\,(e^x - 1 - x)/x) = (1/\alpha - 1)/L + \alpha(1 - \tau) \quad (34)$$

The first parts of these equations have been obtained previously in the original paper suggesting the AK model.[1] Equations 21, 30, and 34 are used to evaluate three parameters in the AK model.

C. Results

1. Nucleic Bases and Nucleosides

Osmometry results are tabulated in Table 2. For 5-bromouridine $\rho = 0.36$, in accord with the original paper that indicated a smaller dimerization constant than equilibrium constant for subsequent additions.[19] In contrast, a high $\rho = 2.7$ for inosine expresses that a greater constant prevails for dimerization.[20] However, this value is so out of line that it is suspect.

2. Urea Association

Urea illustrates what may be expected from precise osmometry data. In contrast to sucrose and glycerol, the osmotic coefficient of urea decreases with increasing concentration,[11] and it must be undergoing self-association. Alone among the entries in Table 2 it does not undergo stacking but rather association through hydrogen bonds. The equilibrium constant for association is so weak that at low concentrations only dimer is present. A cyclic dimer, analogous to acetic acid dimer, may possess a special stability. We are able to use eqs 32 and 34 to test for higher n-mers at higher concentrations. Osmotic coefficients are very accurately known to nearly saturated solutions at 20 molal urea and are judged reliable to 0.001 in Φ.[11] No model mentioned here fits the results with such accuracy to 20 m, so we use the results to a still very concentrated 13.0 m or 8.2 M, the urea concentration used in many protein studies. The dimer model still does not fit, suggesting occurrence of higher n-mers. As may be seen from Table 2, both the EK and AK models provide excellent fits over 34 data points to 13 molal to 0.00065 and 0.00062 Φ units, respectively, better than the precision of the data. We conclude that in more concentrated urea solutions more than dimers occur and that dimers are not especially favored as $\rho < 1$ and $\tau < {}^2/_3$.

Equilibrium constants for dimer formation in urea in the two models are in excellent agreement: $K_2 = \rho K_E = 0.052(1)$ in the EK model and $K_2 = \tau K_A/2 = 0.050(1)$ in the AK model, for an average value of $K_2 = 0.051(1)\ m^{-1}$. An equilibrium constant of 0.041 m^{-1} was previously deduced from a more limited concentration range of the same data without considering nonideality.[18] As expected, omission of nonideality results in too low a value for the equilibrium constant. Owing to the weak association, even in an 8 M solution more than half of the urea molecules still occur as monomers, about one-quarter as dimers, and the remainder as higher n-mers. Despite the very weak association, the highly reliable data and this analysis reveal a wealth of information on the nature of urea solutions.

3. Acetic Acid Dimerization

It is interesting to compare the equilibrium constant in aqueous solutions for dimer formation in

Table 2. Osmotic Coefficient Self-Association Results

			equal *K* model (EK)				attenuated *K* model (AK)			
solute	points[a]	range[b]	K_E	ρ	B	σ^c	K_A	τ	B	σ^c
purine[d]	19	1.1 *m*	2.38(5)	0.94(2)	0.70(7)	2.7	8.2(3)	0.50(3)	0.17(7)	2.5
							10.2(2)	0.35(2)	0.7 set	5.6
6-methylpurine[e]	19	0.70 *m*	6.8(2)	0.98(6)	1.0(2)	4.1	24.3(6)	0.50(3)	−0.09(9)	4.0
							29.7(4)	0.32(2)	0.7 set	6.9
inosine[f]	7	0.11 *m*	1.4(4)	2.7(10)	0.7 set	5.7	5.3(14)	1.3(5)	0.7 set	5.8
deoxyadenosine[f]	6	0.06 *m*	11(2)	1.2(3)	0.7 set	10	38(7)	0.7(2)	0.7 set	11
N^6,N^6-dimethyladenosine[g]	5	0.20 *m*	24(1)	0.8(1)	0.7 set	9.3	120(7)	0.18(5)	0.7 set	15
ethenoadenosine[h]	8	0.08 *m*	21(1)	0.8(2)	0.7 set	15	90(8)	0.27(8)	0.7 set	15
cytidine[d]	18	0.71 *m*	1.4(2)	0.54(8)	0.6(3)	4.2	4.3(7)	0.32(6)	0.2(2)	4.1
			1.51(2)	0.49(2)	0.7 set	4.1	6.1(2)	0.21(2)	0.7 set	4.9
uridine[d]	22	0.71 *m*	0.7(3)	0.8(3)	0.2(5)	7.9	2.2(3)	0.52(6)	0.0(2)	7.9
			1.07(4)	0.60(5)	0.7 set	7.9	4.1(2)	0.29(3)	0.7 set	8.4
5-bromouridine[e]	22	0.42 *m*	3.3(2)	0.36(3)	1.5(2)	4.2	10.8(11)	0.19(3)	0.7(2)	4.2
urea[i]	34	13.0 *m*	0.060(1)	0.86(2)	0.357(7)	0.65	0.177(3)	0.56(1)	0.297(4)	0.62

[a] Number of experimental points. [b] Upper limit of reported concentration range in molality (*m*). [c] Standard deviation from nonlinear least-squares fit in milliosmotic coefficients. Numbers in parentheses for three determined parameters indicate one standard deviation in last digit(s). [d] From ref 36. [e] From ref 19. [f] From ref 20. [g] From ref 45. [h] From ref 49. [i] From ref 11. (13.0 molal corresponds to 8.2 M urea.)

urea with that for acetic acid, which is well known to dimerize in the vapor phase. In the classic study of the determination of the ionization constant of acetic acid from conductance measurements, the authors remark that they are unable to account for the decrease in the value of the ionization constant that occurs above 0.01 M acetic acid.[21] It was subsequently suggested that this decrease may be accounted for by dimerization of neutral acetic acid, and an equilibrium constant estimated to account for the decrease.[22] The treatment involves several mathematical approximations and uses only the Debye−Huckel limiting law. In what follows we provide a more general treatment without the approximations, invoking a different setup to reach a very reliable dimerization constant for acetic acid in aqueous solutions.

Neutral acetic acid undergoes two reactions, dimerization and ionization described by two equilibrium constants

$$HA + HA \rightleftharpoons (HA)_2 \qquad K_D = [(HA)_2]/[HA]^2$$

$$HA \rightleftharpoons H^+ + A^- \quad K_a = [H^+][A^-]/([HA] + 2[(HA)_2])$$

where the dimer contributes to the un-ionized forms in the denominator of the concentration constant K_a. The total concentration of carbon in all its forms is given by $C_T = [A^-] + [HA] + 2[(HA)_2]$, and from which we define mole fractions for each of the three species:

$$\alpha_1 = [A^-]/C_T \qquad \alpha_2 = [HA]/C_T \qquad \alpha_3 = 2[(HA)_2]/C_T$$

The three mole fractions sum to unity. Calculation of literature values from $K_a = \alpha_1^2 C_T/(1 - \alpha_1)$ requires that the $2[(HA)_2]$ term appears in the denominator of the definition for K_a above. Straightforward substitution leads to $\alpha_1^2 = (\alpha_2 + \alpha_3)K_A/C_T$ and $\alpha_3 = 2K_D C_T \alpha_2^2$. With the mole fraction sum, the last equation enables a quadratic expression for α_2:

$$\alpha_2 = \frac{-1 + \sqrt{1 + 8K_D C_T(1 - \alpha_1)}}{4K_D C_T}$$

With these substitutions into the equation for α_1^2 we obtain an expression for α_1 in terms of the parameters to be fitted.

The zero ionic strength ionization constant relates to the concentration constant above by $K_a^0 = K_a\gamma_\pm^2$, where $\gamma_\pm$ is the mean ion activity coefficient. Activity coefficients for neutral species are usually considered to be close to unity. The mean ion activity coefficient is given by the extended Debye−Huckel expression with an additional term that is linear in concentration:

$$-\log \gamma_\pm = 0.51\sqrt{\alpha_1 C_T}/(1 + 0.33a\sqrt{\alpha_1 C_T}) - bC_T$$

The distance of closest approach term in the denominator, $0.33a$, cannot be determined reliably, is of minor significance, and the product is set to 2, representative of values for acids. The coefficient of the linear term, *b*, is refined in the nonlinear least-squares analysis and becomes important in including in the fits points at higher concentrations. It includes contributions from activity coefficients of the ions and neutral molecules and from any deficiency in assignment of the $0.33a$ term.

The quantity most closely related to the experimentally observable conductances is α_1, which is given by the ratio of the equivalent conductance at a given concentration divided by the equivalent conductance of completely ionized acid at the same concentration. Therefore, in the nonlinear least-squares analysis we minimize the error in α_1, the mole fraction of acetate ion.

The above equations provide outstanding fits to the high-quality experimental data obtained at 25 °C. For 18 points to 231 mM acetic acid we find $K_a^0 = 1.7527(5) \times 10^{-5}$ M ($pK_a = 4.7563(1)$), $K_D = 0.192(2)$ M^{-1}, and $b = 0.061(1)$ M^{-1}. In addition to finding a very accurate value for the dimerization constant, K_D, we also obtain the most accurate value of the zero ionic strength ionization constant, K_a^0, from conductances. Deleting the higher concentration points does not alter the K_a^0 value, indicating the constant is consistent with all points, not just those involved in an extrapolation to infinite dilution. The literature

value obtained from extrapolation through points at high dilutions is 1.753×10^{-5} M.[21]

For such a weak reaction the dimerization constant value of $K_D = 0.192(2)$ M^{-1} is well defined and finds support from a value of 0.185(11) M^{-1} derived from the entirely different technique of vapor pressure measurements,[23] both values applying to acetic acid in aqueous solutions. This concordant pair of results provides strong evidence supporting acetic acid dimer formation in aqueous solutions. (A greater value inferred from freezing point depressions is unreliable because not all factors have been included and the points do not extrapolate through the origin.[24]) From results on the vapor phase of acetic acid,[23] we calculate $K_D^v = 5.7 \times 10^4$ M^{-1}, demonstrating that the competition by water in aqueous solutions reduces the dimerization constant by 3×10^5, or the energy equivalent of about two hydrogen bonds.

Owing to the rapidly decreasing mole fraction of acetate ion as the concentration of acetic acid increases from zero and dimer formation becomes significant, a maximum occurs in the mole fraction of un-ionized, monomeric HA at $C_T = 0.035$ M and $\alpha_2 = 0.964$. The weight fraction of dimer $(HA)_2$ surpasses that of ionized A^- at $C_T = 0.053$ M with $\alpha_1 = \alpha_3 = 0.019$. In 0.1 M acetic acid the mole fractions are as follows: A^-, 0.014; HA, 0.951; and $(HA)_2$, 0.035. If the above constants may be applied to more concentrated solutions, the mole fraction of molecules in the dimer surpasses the mole fraction of monomer at 5.3 M acetic acid.

The equilibrium constant for dimer formation in aqueous solutions is almost four times greater for acetic acid than for urea. To the extent that dimer formation depends on the push−pull of hydrogen bond donation and acceptance, a combination of an acidic hydrogen and basic site should aid dimer formation. Typically in amides there is a 16 log unit difference between acidic and basic strengths,[25] while in carboxylic acids the difference is about 11 log units.[26] Alternatively, while amides are about 5 log units more basic than carboxylic acids, the latter are about 10 log units more acidic than amides. The net difference favors greater dimer formation in acetic acid, as observed.

VI. Heats of Dilution

From calorimetrically determined heats of dilution one may obtain primarily the enthalpy change upon dissociation and secondarily the equilibrium constant.[27] All heat effects upon dilution are assumed due to dissociation. The heat of infinite dilution is given by

$$\phi_L C_T = [A_2](\Delta H_2^\circ) + [A_3](\Delta H_2^\circ + \Delta H_3^\circ) + ...$$

If all the heat effects for monomer loss are equal, independent of chain length, then $\Delta H^\circ = \Delta H_2^\circ = \Delta H_3^\circ = ...$, and from eq 1 we find

$$\phi_L C_T = (\Delta H^\circ)([A_2] + 2[A_3] + 3[A_4] + ...) = (\Delta H^\circ)[A](K_2[A] + 2K_2K_3[A]^2 + 3K_2K_3K_4[A]^3 + ...)$$

If only monomer and dimer exist the equation becomes

$$\phi_L C_T = (\Delta H^\circ)K_2[A]^2 = (\Delta H^\circ)\alpha L[A]$$

Recalling that the mole fraction of monomer $\alpha = [A]/C_T$, we now deduce for the dimer only model the relative molal enthalpy

$$\phi_L = \alpha^2 L(\Delta H^\circ)$$

This equation is used with eq 6 for α.

If all the equilibrium constants are equal, we may apply the EK model, and application of eqs 7 and 8 yields

$$\phi_L C_T = (\Delta H^\circ)[A]x(1 + 2x + 3x^2 + ...) = (\Delta H^\circ)[A]L$$

the last equality resulting from eq 9. We now deduce for the relative molal enthalpy

$$\phi_L = \alpha L(\Delta H^\circ)$$

The last equation may be used with eq 11 for α to obtain

$$\phi_L = (\Delta H^\circ)(2L + 1 - \sqrt{4L + 1})/2L$$

This equation places only the most error-prone observable on the left and is the desired form for nonlinear least squares. A linearized form of this equation has been used to estimate ΔH° in the literature studies.

We also go further and allow the dimerization constant to be different from subsequent equilibrium constants and using the same arguments as above with eq 12 instead of eq 7 obtain

$$\phi_L = \alpha\rho L(\Delta H^\circ)$$

This equation is used with the cubic eq 14 for α in the nonlinear least-squares treatment to solve for ΔH°, K_E, and ρ.

Since the initial argument for the attenuated K (AK) model assumed equal enthalpies, it is unfortunate that it does not seem possible to set up heats of dilution in this model along the lines of the other physical methods.

Accurate heat of dilution experiments are available for urea. The relative molal enthalpy is given to high concentration by an equation that is fourth degree in molality.[28] This equation was used to generate ϕ_L values for the same 34 points to 13.0 m used in the osmotic study of section V. From the heats the nonlinear least-squares analysis yields for the association $\Delta H^\circ = -2.12(4)$ kcal/mol, $K_E = 0.052(1)$ m^{-1}, and $\rho = 0.78(3)$. The overall fit is not as good as that for the osmotic coefficient listed at the bottom of Table 2, but the good agreement between the widely different methods serves to support the analyses and similar conclusions ($\rho < 1$). The heat change compares closely with that of −2.09 kcal/mol evaluated from the same data at lower concentrations by another method.[18]

A similar nonlinear least-squares analysis was performed on several purines[29] and pyrimidines[30] for which heats of dilution have been published. However, owing to relatively high scatter, data for most of the compounds do not refine well. The difficulty

of determining K_E from the heat data has prompted the thermochemists to combine calorimetric and osmotic studies. However, we only analyze the calorimetric results internally.

Consistent data have been published for purine and several 9-alkyl-substituted purines.[31] For 7 points to 1.0 M purine we calculate $K_E = 3.39(3)$ M^{-1}, $\rho = 0.86(4)$, and $\Delta H° = -3.6(1)$ kcal/mol, leading to an entropy change of $\Delta S° = -9.5$ cal/(deg mol). With ρ assumed as 1.00, the original paper reported $K_E = 3.1(1)$, $\Delta H° = -3.21(5)$, and $\Delta S° = -8.5$ cal/(deg mol).

For 8 points to 0.50 *m* 6-methylpurine[27] we calculate $K_E = 8.61(4)$ m^{-1}, $\rho = 0.96(2)$, and $\Delta H° = -5.8(1)$ kcal/mol. (The dimer only model does not fit the data.) If we assume as the authors $\rho = 1$, then $K_E = 8.50(8)$ m^{-1} and $\Delta H° = -5.63(2)$ kcal/mol, nearly identical to the literature value for the same data. For stacking in 6-methylpurine the enthalpy change of $\Delta H = -5.8$ kcal/mol when combined with the free energy change leads to an entropy change of $\Delta S = -15$ cal/(deg mol).

For both purine and 6-methylpurine the values of K_E derived from heat measurements are somewhat greater than those found by other physical methods and reviewed in this article. The thermodynamic quantities are typical of those found for stacking interactions. Stacking in other compounds yields similar values with negative enthalpy and entropy changes.[29−32] Stacking is enthalpy driven; the entropy change is unfavorable. These results are inconsistent with stacking being attributed solely to hydrophobic interactions, which display small positive enthalpy and large negative entropy changes.[33]

VII. Partition Coefficients: Caffeine

The partitioning of caffeine between water and an immiscible organic solvent more than doubles with caffeine concentration as stacks form in water but not in the organic solvent. It is easy to apply the EK and AK models to partitioning. The partition coefficient, PC, defined as the total concentration of caffeine in the aqueous phase to the concentration in the organic phase, is given by

$$\mathrm{PC} = \frac{C_{aq}}{C_{org}} = \frac{[A_{aq}] + 2[A_2] + 3[A_3] + ...}{[A_{org}]} = \frac{[A_{aq}](1 + 2K_2[A_{aq}] + 3K_2K_3[A_{aq}]^2 + ...)}{[A_{org}]}$$

where we have applied eq 1. We also define an equilibrium constant for the mole ratio of monomer in aqueous to organic phase, $K_m = [A_{aq}]/[A_{org}]$.

For the case where only monomer and dimer exist the above equation becomes

$$\mathrm{PC} = K_m(1 + 2K_2[A_{aq}]) = K_m(1 + 2\alpha L) = K_m/\alpha$$

the last equality following from eq 5 and where α is given by eq 6.

In the EK model we apply eqs 8, 9, and 12 to obtain

$$\mathrm{PC} = K_m(1 + \rho(2x + 3x^2 + ...)) = K_m\frac{1 - x(2 - x)(1 - \rho)}{(1 - x)^2} = K_m/\alpha$$

This equation is used with eq 14 for α in the nonlinear least-squares analysis in the EK model.

In the AK model we invoke eqs 17−19 to derive

$$\mathrm{PC} = K_m(1 + \tau(x + x^2/2 + x^3/6 + ...)) = K_m(1 + \tau(e^x - 1)) = K_m/\alpha$$

This equation is used with eq 21 for α in the nonlinear least-squares analysis in the AK model.

In the dimer only case and both models $\alpha = K_m/\mathrm{PC}$, the mole fraction monomer in the aqueous phase at any concentration is given by the ratio of the constant K_m to the experimental PC.

Partitioning of caffeine (1,3,7-trimethyl-2,6-dioxopurine) between water and an immiscible organic solvent mixture has been reported graphically in 35 points to 0.13 M total caffeine at 30 °C in the aqueous phase.[34] From a blown up graph to interpolate the points, a nonlinear least-squares analysis gives in the EK model $K_E = 7.7(2)$ M^{-1}, $\rho = 1.16(6)$, and $K_m = 23.1(1)$, and in the AK model $K_A = 29(1)$ M^{-1}, $\tau = 0.58(5)$, and $K_m = 23.1(1)$. The fit with the dimer only model is unsatisfactory and it may be excluded; higher *n*-mers occur.

The EK values may be compared with those calculated from 8 points to 0.10 *m* at 25 °C from heats of dilution data by the protocol of section VI to give $K_E = 17(2)$ m^{-1} and $\rho = 1.0(5)$. From our more self-contained and sophisticated calculation, the enthalpy of stacking of caffeine is $\Delta H = -2.4(8)$ kcal/mol, significantly less than the literature value of −3.4 kcal/mol.[29] For comparison below, at 30 °C the equilibrium constant is reduced to $K_E = 16$ m^{-1}. Again, the dimer only model does not yield a fit.

We have also blown up a graph that yields 11 points to 0.10 M caffeine in water at 30 °C of upfield chemical shifts for four protons.[35] By the simultaneous nonlinear least-squares analyses described in section IX we find in the EK model $K_E = 13(3)$ M^{-1} and $\rho = 1.0(2)$, and in the AK model $K_A = 35(5)$ M^{-1} and $\tau = 0.6(1)$.

In summary, for three very different experimental methods we find for caffeine at 30 °C the ranges $7.7 < K_E < 16$ and $1.0 < \rho < 1.2$ in the EK model, and $29 < K_A < 35$ and $\tau = 0.6$ in the AK model. Only a factor of about 2 in the range for K_E spoils the agreement in a case the accuracy of which is limited by the solubility of caffeine in water. Incorporation of caffeine into purine stacks is discussed in section X.

VIII. Comparisons between Two Methods and Two Models

Owing to the differences in set up, the equilibrium constants obtained in the equal *K* (EK) and attenuated *K* (AK) models are not directly comparable. However, we may compare the EK and AK models for low amounts of stacking by relating the equilibrium constants for dimer and trimer formation.

From eqs 12 and 19 we have for dimer formation $\rho K_E = K_2 = \tau K_A/2$ and for trimer formation $K_E = K_3 = K_A/3$ that together yield

$$K_A \approx 3K_E \quad \text{and} \quad \tau \approx 2\rho/3 \tag{35}$$

These relations are approximately obeyed in Tables 1 and 2 (and later Tables 3 and 4), the more so the fewer higher *n*-mers.

There is a persistent tendency in Tables 1 and 2 (and later in Table 4) for $\tau < 0.67$. In the AK model from eq 19 we obtain $K_2/K_3 = 3\tau/2$. Values of $\tau < 0.67$ imply that $K_2 < K_3$, consistent with an additional entropy loss in dimer compared to trimer formation as the former results in loss of orientational freedom in two monomers. Values of $\tau = 0.67$ in the AK model compare with $\rho = 1.00$ in the EK model. Thus a major failing of the simple attenuated AK model with $\tau = 1$ is that $K_2 = 1.5K_3$, when the evidence indicates that for most cases $K_2 < K_3$, and $\tau < 0.67$ in the more complete AK model makes this adjustment. The attenuated model attenuates too soon. Thus it seems necessary to include τ in the AK model. If the precision of the data does not allow τ as a parameter to be determined, it should be set to $\leq 2/3$, not to 1.00.

Within the EK model there is excellent agreement between sedimentation and osmometry results for purine and cytidine. With a lesser concentration range and number of points, agreement is less satisfactory for inosine and deoxyadenosine. When the concentration range is limited and/or the complexes are weak, it is appropriate to compare the products given above as together they determine the dimerization constant K_2. On this basis there is improved agreement between the sedimentation and osmometry results within both the EK and AK models.

This study appears the first to include a nonideal term in an osmotic analysis for stacking. Compared to values of constants evaluated without the nonideal term,[1] our values of K_E and K_A are greater and the values of ρ and τ are less. In both the EK and AK models the determined value for the second virial coefficient, *B*, is smaller for osmometry than for sedimentation, and takes on unreasonably low value in several AK cases. In the attenuated AK model the ever-decreasing equilibrium constants produce the same mathematical effect as the osmotic nonideal term, and the nonlinear least-squares refinement does not distinguish the two in any but the most precise cases. It is probably better to set a fixed *B* value and refine for the other two parameters. This has been done for several cases in Table 2.

The original osmotic coefficient studies reported a decrease in the value of an association constant with increasing concentration of purine[36] and 6-methylpurine.[19] Without the nonideal *B* term the calculated ρ values are 1.31(5) and 1.40(5), respectively, accounting quantitatively for the observation. Likewise, these $\rho > 1$ values were cited as an argument for the attenuated (AK) model over the equal (EK) model.[1] Inclusion of the nonideal term yields ρ values near unity, however, and the Table 2 results indicate that the equilibrium constant decrease and high ρ values are only apparent and that both effects are accommodated wholly by the nonideal *B* term. Thus it is important to include the nonideal term with osmotic coefficient data.

We are now in a position to compare the results of sedimentation and osmometry analyses with those deduced by an entirely different criterion from chemical shifts in nuclear magnetic resonance spectroscopy.

IX. NMR Chemical Shifts

It has long been recognized that stacking of nucleic bases results in upfield shifts in the nuclear magnetic resonance spectra. NMR chemical shifts offer at least two strong points as the method of choice in evaluating stacking parameters. First they may be obtained with greater ease and precision than observables in other techniques. Second, as individual nuclei may be studied, with an underlying theory chemical shifts offer the possibility of providing a microscopic picture of stacking.

Owing to rapid exchange in self-association reactions, chemical shifts in nuclear magnetic resonance spectroscopy represent time averages over all species present. In addition to NMR chemical shifts, application may be made to intensities in absorption and circular dichroism spectra (Section IX.D.12).[5,37] The approach is applicable to any property wherein the observable is a weighted average over the species present.

A. Dimer

The dimer represents an important special case which will be compared with EK stacking below. There are only two solute species: monomer with mole fraction α and dimer with mole fraction λ with $\alpha + \lambda = 1$. The weighted average property of the system is then given by

$$P = \alpha P_\alpha + \lambda P_\lambda = \alpha P_\alpha + (1 - \alpha)P_\lambda$$

where P_α and P_λ are the values of the property for monomer and dimer molecules. Rearrangement yields

$$P - P_\alpha = (P_\lambda - P_\alpha)(1 - \alpha) = (P_\lambda - P_\alpha)(1 + (1 - \sqrt{8L + 1})/4L) \tag{36}$$

the last equality resulting from substitution of eq 6. Thus the chemical shift difference between observed and monomer resonances is a function of the chemical shift difference between dimer and monomer and the product $L = K_2C_T$.[38] We also have for the mole fraction monomer, $\alpha = (P_\lambda - P)/(P_\lambda - P_\alpha)$. Knowing the limiting values, one may calculate a dimer equilibrium constant at any point from

$$L = K_2C_T = (1 - \alpha)/2\alpha^2 = (P - P_\alpha)(P_\lambda - P_\alpha)/2(P_\lambda - P)^2 \tag{36a}$$

When the property *P* is midway between the limits P_α and P_λ, we have $L = 1$, $K_2 = 1/C_T$, and $\alpha = 0.50$. At the first $1/6$ of the transformation in *P* we find $L = 0.120$ and at the $5/6$ mark, $L = 15.0$ for a 125-fold (2.10 log units) concentration range over the middle

$^2/_3$ of the transformation, with the halfway point in P at $L = 1.00$ nearer the L value at the $^1/_6$ point.

B. Nearest Neighbors

In the usual isodesmic model for stacking we assign α as the mole fraction of monomer, λ as the mole fraction of molecules at the ends of the stack, and ξ as the mole fraction of molecules within a stack. Then the observed property, P, such as the chemical shift is given by

$$P = \alpha P_\alpha + \lambda P_\lambda + \xi P_\xi \tag{37}$$

where $P\alpha$, P_λ, and P_ξ are the values of the property for monomer, molecule at the end of a stack, and molecule within a stack, respectively. Because there are three parameters to be determined, it is usual to reduce it to two by assuming that the chemical shift of a molecule at the end of a vertical stack is given by the average of monomer and interior molecule shifts: $P_\lambda = (P_\alpha + P_\xi)/2$. This assumption sets $f = 0.50$ in the following equivalent equations:

$$P_\lambda = (1 - f)P_\alpha + fP_\xi \qquad P_\lambda - P_\alpha = f(P_\xi - P_\alpha) \tag{38}$$

The usual setting of $f = 0.50$ rests on the assumption that a molecule at the end of a stack experiences half the effect of one inside a stack. However, we wish to allow P_λ to slide along the P_α to P_ξ axis and seek to find the best nonlinear least-squares fit of $0 < f < 1$ in eq 38. If $f = 0$, the value of the property at the end of the stack is equal to that of the monomer, and if $f = 1$ to that of an internal stack molecule. Since their environments are not the same, different nuclei in the same molecule may exhibit different f values. We now need to determine three parameters: P_α, P_ξ, and f.

The mole fraction of monomer molecules is given by the equations for α in section II. For the mole fraction of molecules at the end of a stack we have

$$\lambda = (2[A_2] + 2[A_3] + 2[A_4] + 2[A_5] + ... + 2[A_i])/C_T \tag{39}$$

and for molecules within a stack

$$\xi = ([A_3] + 2[A_4] + 3[A_5] + 4[A_6] + ... + (i - 2)[A_i])/C_T \tag{40}$$

The three mole fractions sum to unity: $\alpha + \lambda + \xi = 1$.

1. EK Model

The mole fraction of monomer, α, is given by eq 14. Combination of eqs 8, 12, and 39 leads to the mole fraction of molecules at stack ends.

$$\lambda = 2\rho\alpha x/(1 - x) \tag{41}$$

Combination of eqs 8, 12, and 40 leads to the mole fraction of molecules in a stack interior.

$$\xi = \rho\alpha x^2/(1 - x)^2 \tag{42}$$

Equations 14, 37, 38, 41, and 42 are used in the nonlinear least-squares analysis for each nucleus.

The weight fraction of the molecules at the end of a stack, λ, passes through a maximum as the ends grow in prominence at the early stages of stack formation and become a decreasing feature as the stacks lengthen. We find the maximum in λ occurs at the total concentration C_T given by $L = K_E C_T = 2$ with a value of $\lambda_{max} = \sqrt{\rho}/(1 + \sqrt{\rho})$. Only the equilibrium constant K_E determines the concentration at which the maximum occurs, and only ρ determines the value of λ_{max}. Also at the maximum we find $\alpha_{max} = \xi_{max} = [2(1 + \sqrt{\rho})]^{-1}$. At $\rho = 1$ these equations reduce to $\lambda_{max} = 0.50$, and $\alpha_{max} = \xi_{max} = 0.25$.

The usual treatment of NMR chemical shift results is to assume $\rho = 1$, $f = 0.5$ in eq 38, and with eqs 11, 37, 41, and 42, after algebraic manipulation, obtain

$$P - P_\alpha = (P_\xi - P_\alpha)\alpha L = (P_\xi - P_\alpha)(1 + (1 - \sqrt{4L + 1})/2L) \tag{43}$$

Equation 43 is of the same form as the dimer eq 36 with $P_\xi = P_\lambda$ and $K_E = 2K_2$. Thus in this approximation the isodesmic stack and the dimer become formally indistinguishable.[38] A high $P_\xi - P_\alpha$ value of > 1.0 ppm for protons suggests stacking beyond dimers, but lesser values do not usually imply only dimers. Knowing the limiting values one may estimate the equilibrium constant from

$$L = K_E C_T = \alpha L/(1 - \alpha L)^2 = (1 - \sqrt{\alpha})/\alpha = (P - P_\alpha)(P_\xi - P_\alpha)/(P_\xi - P)^2 \tag{43a}$$

When the property P is midway between the limits P_α and P_ξ, we have $L = 2$, $K_E = 2/C_T$, and $\alpha = 0.25$.

2. AK Model

The mole fraction of monomer is given by eq 21. Combination of eqs 17, 19, and 39 gives for stack end mole fraction

$$\lambda = 2\tau(e^x - 1 - x)/L \tag{44}$$

and combination of 17, 19, and 40 give for stack interior mole fraction

$$\xi = 1 - \alpha - \lambda \tag{45}$$

Equations 21, 37, 38, 44, and 45 are used in the nonlinear least-squares analysis.

C. Next Nearest Neighbors

The treatment in the previous section assumed that only nearest-neighbor interactions produce chemical shifts. In an entirely new development we consider also the interaction of next nearest neighbors in vertical stacks. This treatment is not a refinement but a more detailed alternative approach to the one above.

Let the nearest-neighbor effect producing an NMR chemical shift be n and the next-nearest-neighbor effect be t, with their ratio $r = t/n$. A theoretical paper for proton chemical shifts in the nucleic bases estimates that $r \approx 0.15$.[39] A ratio this high suggests that the next nearest neighbor is too important to

be ignored. Since their environments are not the same, different nuclei in the same molecule may display different r values.

We distinguish five categories of molecules in a vertical stack designated by their mole fractions β, γ, δ, ϵ, and ζ, that together with the mole fraction of monomer, α, sum to unity. In a dimer there are only nearest-neighbor interactions, n, for both molecules (δ), $2n$ interactions for a molecule in the center of a trimer (γ), $n + t$ interactions for molecules at ends of stacks of three or more molecules (β), $2n + t$ interactions for penultimate molecules in all stacks of four or more molecules (ϵ), and $2n + 2t$ interactions for all other interior molecules beginning with the central one in a five molecule stack (ζ). (See the bank of eqs 49 below for more description.)

We define the chemical shifts from monomer to a stacked molecule as follows:

$$\Delta\beta = P_\beta - P_\alpha = n + t = n\,(1 + r) \qquad (46)$$

$$\Delta\gamma = P_\gamma - P_\alpha = 2n$$

$$\Delta\delta = P_\delta - P_\alpha = n$$

$$\Delta\epsilon = P_\epsilon - P_\alpha = 2n + t = n(2 + r)$$

$$\Delta\zeta = P_\zeta - P_\alpha = 2n + 2t = 2n\,(1 + r)$$

Comparing each of the first four shifts to the greatest shift $\Delta\zeta$ for wholly interior molecules we derive

$$P_\beta = (P_\alpha + P_\zeta)/2 \qquad P_\gamma = (rP_\alpha + P_\zeta)/(1 + r)$$
$$P_\delta = \frac{(1 + 2r)P_\alpha + P_\zeta}{2(1 + r)} \qquad \frac{P_\epsilon = rP_\alpha + (2 + r)P_\zeta}{2(1 + r)} \qquad (47)$$

The set of four equations provide chemical shifts of four categories of vertical stack molecules in terms of the chemical shifts of monomer and wholly interior molecules. They are used in combination with the following equation for the observed chemical shift.

$$P = \alpha P_\alpha + \beta P_\beta + \gamma P_\gamma + \delta P_\delta + \epsilon P_\epsilon + \zeta P_\zeta \quad (48)$$

From the definition of the molecule settings in a vertical stack we may write the following equations defining each kind of mole fraction:

$$\delta = 2[\mathrm{A}_2]/C_\mathrm{T} \qquad (49)$$

$$\beta = (2[\mathrm{A}_3] + 2[\mathrm{A}_4] + 2[\mathrm{A}_5] + 2[\mathrm{A}_6] + 2[\mathrm{A}_7] + \ldots)/C_\mathrm{T}$$

$$\gamma = [\mathrm{A}_3]/C_\mathrm{T}$$

$$\epsilon = (2[\mathrm{A}_4] + 2[\mathrm{A}_5] + 2[\mathrm{A}_6] + 2[\mathrm{A}_7] + \ldots)/C_\mathrm{T}$$

$$\zeta = ([\mathrm{A}_5] + 2[\mathrm{A}_6] + 3[\mathrm{A}_7] + \ldots)/C_\mathrm{T}$$

Note how each stack length is fully accounted for; for example in the six molecule stack, A_6, pairs of molecules are distributed among three of the mole fractions.

Remaining is to derive expressions for mole fractions in terms of the monomer mole fraction α.

1. EK Model

Combination of eqs 8, 12, and 49 leads to

$$\beta = 2\rho\alpha x^2/(1 - x) \qquad \gamma = \rho\alpha x^2 \qquad \delta = 2\rho\alpha x$$
$$\epsilon = 2\rho\alpha x^3/(1 - x) \qquad \zeta = \rho\alpha x^4/(1 - x)^2 \qquad (50)$$

Equations 14, 47, 48, and 50 are used in the nonlinear least-squares analysis for each nucleus.

2. AK Model

Combination of eqs 17, 19, and 49 leads to

$$\beta = 2\tau\alpha(\mathrm{e}^x - 1 - x - x^2/2)/x \qquad \gamma = \tau\alpha x^2/6$$
$$\delta = \tau\alpha x \qquad \epsilon = 2\tau\alpha(\mathrm{e}^x - 1 - x - x^2/2 - x^3/6)/x$$
$$\zeta = 1 - \alpha - \beta - \gamma - \delta - \epsilon \qquad (51)$$

Equations 21, 47, 48, and 51 are used in the nonlinear least-squares analysis.

It is instructive to compare the two fundamentally different alternative treatments in sections IX.B and IX.C, the latter considerably more elaborate. In both the equal K (EK) and attenuated K (AK) models, by equating the limiting chemical shifts for molecules in the interior of a stack, $P_\xi = P_\zeta$, it may be shown that the relationship between f and r is given by

$$2f = 1/(1 + r) \qquad (52)$$

Thus the f treatment, through eq 52, also allows for next-nearest-neighbor interactions. For no next-nearest-neighbor effects $r = 0.00$ and $f = 0.50$. As next-nearest-neighbor effects increase, r increases and f decreases: for $r = 0.25$, $f = 0.40$, and for $r = 0.50$, $f = 0.33$. Thus the decrease in f is not as great as the increase in r. Values of $f > 0.50$ (see section IX.D.11) correspond to no positive values of r, and imply a stepped rather than a vertical stack, which demands a new setup in the next-nearest-neighbor treatment. Equation 52 will not always be realized in nonlinear least-squares fits because the minimization may result in differences in other parameters, such as the limiting chemical shift for molecules in the interior of a stack. And there may be other stack particulars that affect f and r in different ways.

D. Results

At the outset we need to recognize that the nonlinear least-squares treatment is being challenged by the large number of parameters to be determined. For a given compound we have two parameters independent of the particular nucleus, either K_E and ρ in the EK model or K_A and τ in the AK model, plus three parameters for each nucleus: f or r depending upon the treatment, and the limiting chemical shifts of the monomer and a molecule in the interior of a stack. The last quantity carries a standard deviation of up to 0.1 ppm because only in optimum (large L) cases is the observed chemical shift at the highest

Table 3. Proton NMR Stacking Parameters[a]

	purine[b]		6-methylpurine[c]	
	EK model	AK model	EK model	AK model
K_E or K_A[d]	$K_E = 2.43(14)$	$K_A = 6.7(3))$	$K_E = 4.1(6)$	$K_A = 12(1)$
ρ or τ	$\rho = 1.05(17)$	$\tau = 0.54(5)$	$\rho = 0.9(2)$	$\tau = 0.3(1)$
f, H2	0.26(6)	0.29(2)	0.45(6)	0.53(4)
H8	0.22(6)	0.25(3)	0.40(8)	0.48(5)
H6	0.28(6)	0.32(2)		
$P_\xi - P_\alpha$,[e] H2	1.6	1.8	1.2	1.2
H8	1.2	1.4	1.0	1.0
H6	1.8	2.0		
K_E or K_A[d]	$K_E = 2.15(8)$	$K_A = 5.7(3)$	$K_E = 4.1(6)$	$K_A = 12(1)$
ρ or τ	$\rho = 0.97(6)$	$\tau = 0.58(5)$	$\rho = 0.9(2)$	$\tau = 0.3(1)$
r, H2	0.44(8)	0.41(9)	0.2(2)	−0.04(9)
H8	0.58(13)	0.60(14)	0.2(3)	0.08(15)
H6	0.38(7)	0.34(8)		
$P_\zeta - P_\alpha$,[e] H2	1.5	1.7	1.2	1.2
H8	1.2	1.3	1.0	1.0
H6	1.7	1.9		

[a] Above horizontal line near center is nearest neighbor (*f*) treatment of section IX.B and below line next nearest neighbor (*r*) treatment of section IX.C. [b] Calculated for 20 points for each nucleus to 1.04 molal with apparent σ of 0.012 ppm from a table of data described in ref 40. [c] Calculated for 9 points for each nucleus to 0.94 molal with apparent σ of 0.017 ppm from a blowup of Figure 2 in ref 44. [d] In units of m^{-1}. Numbers in parentheses represent one standard deviation in last digit(s). [e] Upfield shift in ppm from monomer to a molecule in interior of stack. Standard deviations in the difference about 0.1 ppm.

concentration half way to the extrapolated chemical shift of an interior molecule. A special feature of our objective nonlinear least-squares analysis is that it minimizes *simultaneously* the differences between observed and calculated chemical shifts for the several nuclei. Thus for purine with three protons there are $2 + 3 \times 3 = 11$ parameters to be fitted (3.7 per nucleus) and separately with five carbons $2 + 3 \times 5 = 17$ parameters (3.4 per nucleus). Given this flexibility and the consequent shallow minimums it is reassuring that the consistency of the fits is so good. Other values of the parameters may yield fits nearly as good as the minimums reported, but there is too vast an array of good but not minimal fits to explore. Depending upon the concentration units of the original paper, units of the equilibrium constants K_E and K_A are either molar^{-1} or molal^{-1}. Although NMR is the simplest of the methods, precise determination of upfield chemical shifts due to stacking depends upon an appropriate reference. External references do not allow easily for bulk susceptibility corrections and some internal references undergo shifts in the presence of aromatic molecules.[40−42] As it is unaffected by the presence of aromatic molecules, tetramethylammonium ion has been used as an internal reference in several of the cases reported herein.[42] It undergoes a usually insignificant 0.0058 ppm upfield shift per mole ionic strength of added salt.[43]

In order to evaluate convincingly ρ and τ of the two models and especially *f* and *r* of the two treatments we need compounds for which the stacking becomes extensive enough so that at the highest concentrations at least $^1/_3$ of the molecules are in stacks of trimer or higher. Equations 15 and 22 may be used to provide such estimates by evaluating $1 - \alpha - \alpha_2$; values of $K_E C_T(\max) = L(\max) > 1.0$ meet the criterion. When $L(\max) > 2.0$, more than half the molecules reside in stacks of three or more molecules. (The conclusions are almost independent of the ρ value.) The results in Tables 3 and 4 indicate that both purine and 6-methylpurine surpass this criterion; at the highest concentration more than half the molecules reside in stacks of three or more molecules.

1. Purine

Twenty points to 1.04 molal are available for purine[40] and results of a nonlinear least-squares analysis of three protons simultaneously appear in Table 3. Viewing Table 3 vertically we see that within each model the *f* and *r* treatments yieldsimilar values for K_E and ρ on one hand and K_A and τ on the other. As found in Tables 1 and 2, we observe $K_E < K_A$ and $\rho > \tau$. Upfield chemical shifts from monomer to interior molecule are similar for the two models, although slightly greater for the AK model foreshadowing a greater discrepancy in the carbon-13 results reported below.

Purine proton NMR exhibits exceptional behavior that is not repeated by other compounds under review. For purine the calculated values of *f* and *r* for each proton are virtually identical regardless of whether the EK or AK model is used. However, the values of these parameters are significantly different from 0.50 and 0.00, respectively. Especially the high *r* values ranging from 0.3 to 0.6 suggest important next nearest neighbor interactions for protons in purine. Equation 52 is only roughly obeyed but there are large standard deviations in *r*. Setting $f = 0.50$ or $r = 0.00$ results in slightly poorer fits. Since purine proton NMR is the only case for which such high *r* and low *f* values have been found, we need to reserve judgment until the experiments are repeated and the new data processed.

Carbon-13 NMR results for purine[40] supplement those in Table 3 for protons. For 12 points for each nucleus to 1.01 molal five carbon atoms, C2, C4, C5, C6, and C8, were analyzed simultaneously by nonlinear least squares. In both the EK and AK models the values calculated for all five carbons for *f* and *r* came within one standard deviation of $f = 0.50$ and $r = 0.00$, so these parameters were set at these values in subsequent runs. Thus the next-nearest-neighbor

Table 4. NMR Stacking Results[a]

solute	nuclei−points[b]	range[c]	σ[d]	EK model		AK model	
				K_E	ρ	K_A	τ
purine[e]	3−20	1.04 m	0.012	2.3(2)	1.0(1)	6.2(5)	0.56(5)
purine,[e] ^{13}C	5−12	1.01 m	0.036	3.0(3)	1.0(1)	7(2)	0.8(2)
6-methylpurine[f]	2−9	0.94 m	0.017	4.1(6)	0.9(2)	12(1)	0.3(1)
inosine[g]	3−9	0.10 M	0.0010	4.7(9)	0.8 set	12(2)	0.53 set
adenosine[h]	4−10	0.051 M	0.0030	19(4)	0.83(9)	56(9)	0.48(15)
N^6,N^6-dimethyladenosine[i]	4−5	0.20 m	0.0048	29(1)	1.0(1)	70(5)	0.7(1)
1,N^6-ethenoadenosine[j]	5−17	0.081 M	0.0043	12(1)	0.8(1)	37(4)	0.45(7)
5′-H(AMP)$^-$ [k]	3−14	0.40 M	0.0032	3.4(2)	1.03(3)	5.7(15)	1.0(2)
Mg(ADP)$^-$ [l]	3−10	0.40 M	0.0038	9.7(5)	0.58(3)	24(4)	0.37(8)
Mg(ATP)$^{2-}$ [g]	3−10	0.40 M	0.0036	8.5(3)	0.41(2)	24(5)	0.23(6)
ϵ-ATP^{4-} [m]	5−9	0.40 M	0.0033	3.5(1)	0.66(3)	8(1)	0.45(8)
Mg(ϵ-ATP)$^{2-}$ [m]	5−9	0.40 M	0.0028	9.1(6)	0.77(9)	23(2)	0.47(5)
cytidine[g]	3−9	0.50 M	0.0013	1.6(1)	0.9(1)	5(1)	0.5(1)
uridine	3−9	0.50 M	0.0010	1.1(2)	0.9(1)	4(1)	0.6(2)
1,10-phenanthroline[n]	4−12	0.023 M	0.0040	41(4)	0.81(5)	119(23)	0.47(9)

[a] In water for protons except for purine carbon-13 in second row. [b] Number of nuclei processed simultaneously − number of experimental points for each nucleus. [c] Upper limit of reported concentration range in molality (m) or molarity (M). [d] Standard deviation from nonlinear least-squares fit in chemical shift ppm for both EK and AK models. Numbers in parentheses for determined parameters indicate one standard deviation in last digit(s). [e] From ref 40. [f] From blowup of Figure 2 in ref 44. [g] From ref 42. [h] From ref 38. [i] From blowup of Figure 6 in ref 45. [j] From ref 47. [k] From ref 50. [l] From ref 51. [m] From ref 53. [n] From ref 55.

interaction in purine appears to be less for carbon than for protons. In the EK model we find $K_E = 3.0$-(3) and $\rho = 1.0(1)$, and upfield chemical shifts of 1.8, 3.0, 2.2, 2.0, and 1.5 ppm, respectively. In the AK model we calculate $K_A = 7(2)$ and $\tau = 0.8(2)$, and upfield shifts of 2.1, 3.6, 2.6, 2.4, and 1.9 ppm, respectively. All four analyses yield an apparent standard deviation of 0.036 ppm. From the two distinct sets of five chemical shifts we see that the EK and AK models provide different solutions to the carbon-13 data. As shown atop Table 4, the carbon-13-derived equilibrium constant values correspond well to those for protons in both models.

It is striking that for purine the agreement between five disparate physical methods, sedimentation (Table 1), osmometry (Table 2), heats of dilution (section VI), and proton (Table 3) and carbon-13 NMR (Table 4) give such excellent agreement in the EK model for $2.2 < K_E < 3.4$ and $0.86 < \rho < 1.05$. For the AK model the range of agreement is poorer: $5.7 < K_A < 10.4$ and $0.4 < \tau < 0.8$. The EK model agreements are much better than expected, or realized from the literature because the analyses performed here are new.

2. 6-Methylpurine

Results of a nonlinear least-squares analysis of one of the earliest proton NMR studies on stacking[44] are listed in the last two columns of Table 3. In this, typical case calculated values of f and r are almost within one standard deviation of $f = 0.50$ and $r = 0.00$, suggesting that next-nearest-neighbor interactions are insignificant. The f and r treatments yield closely similar results for the equilibrium constants. Both the EK and AK models yield identical upfield shifts from monomer to interior stack molecule.

In a new insight, despite uniformly greater equilibrium constants, 6-methylpurine exhibits lesser upfield shifts from monomer to stack interior molecule than purine. Owing to the stronger stacking, at a given concentration observed shifts are greater for 6-methylpurine than for purine.

For 6-methylpurine the agreement in the EK model among three physical methods, osmometry, heats of dilution (section VI), and proton NMR is good: $4.1 < K_E < 8.5$ and $0.9 < \rho < 1.0$. In the AK model we may only compare the results for osmometry (Table 2) and proton NMR (Table 4) to find K_A values that differ by a factor of 2. The poorer agreement among values in the AK model agrees with the perception that the nonlinear least-squares refinement works less effectively with the equations for this model.

3. Inosine

Solubility limits the maximum $L \approx 0.4$ so that even in the most concentrated solutions only 15% of the molecules appear in stacks of three or more molecules. Since they are indeterminate in this case, we set $f = 0.50$ and $r = 0.00$, making the treatments equivalent, and also $\rho = 0.80$ and its approximate equivalent, $\tau = 0.53$. Data are available for three nuclei: H2, H8, and H1′ on the ribose.[42] Results of the simultaneous nonlinear least-squares analysis appear in Table 4. The upfield chemical shifts from monomer to molecule within a stack are slightly different in the two models: In EK 0.29, 0.25, and 0.26 ppm, and in AK 0.04 ppm greater. These shifts are among the smallest observed in a purine. In the original paper with $f = 0.00$ and $\rho = 1.00$, in the EK model the average $K_E = 3.3$ from runs on individual nuclei, with 0.04 ppm greater upfield chemical shifts. The ρK_E product is similar in the two calculations (3.8 and 3.3).

Comparison of the inosine equilibrium parameters among three physical methods shows good agreement between sedimentation (Table 1) and proton NMR (Table 4) with $2.6 < K_E < 4.7$ and $\rho = 0.8$ in the EK model, and $9 < K_A < 12$ and $0.4 < \tau < 0.53$(set) in the AK model. The results from osmometry in Table 2 fall well outside of these ranges.

4. Adenosine

Of limited solubility, data are available for 10 points to 0.051 M,[38] where $L \approx 1.0$ and about $^1/_3$ of

the molecules exist as trimers or higher *n*-mers at the highest concentration. The most general fit yields values of $f = 0.50$ and $r = 0.00$ within one standard deviation of these figures, so they are set in subsequent runs. Thus the prediction from theory that $r = 0.16$ for adenine[39] seems too high. Nonlinear least-squares fits performed simultaneously on four protons, H2 and H8 on the base and H1′ and H2′ on the ribose gave excellent fits with the results listed in Table 4. In the EK model the upfield chemical shifts of monomer to interior stack molecule are 0.46(5), 0.25(3), 0.21(3), and 0.17(3) ppm, respectively. In the original paper which assumed $\rho = 1.00$, calculations on individual nuclei gave an average $K_E = 15 \pm 2$ and upfield chemical shifts 11% greater than those above. The ρK_E product is virtually identical in the two calculations. In the AK model the upfield shifts are slightly greater, 0.51(4), 0.28(3), 0.23(2), and 0.19(2) ppm, respectively. The relations of eq 35 comparing the EK and AK models are well-observed. Although both models yield the same apparent standard deviation, the EK model handles better with a more stable determinant.

The stacking parameter values derived for adenosine from NMR show excellent agreement with those for deoxyadenosine determined from sedimentation (Table 1) and osmometry (Table 2). Over the three experimental techniques the range of values is only $11 < K_E < 19$ and $0.83 < \rho < 1.2$ in the EK model and $38 < K_A < 56$ and $0.4 < \tau < 0.7$ in the AK model. The close agreement among the three techniques supports once again their validity and the insignificance of the 2′-oxygen for stacking.

5. N^6,N^6-Dimethyladenosine

Ring substitution and especially methylation strongly increases the tendency to stack. Although only 5 points are available to 0.20 *m*, this methylated adenosine stacks so strongly that more than half the molecules exist as trimers or higher *n*-mers at the higher concentrations.[45] Chemical shift data for the H2, methyl, H8, and H1′ protons were analyzed simultaneously by nonlinear least squares. In the most general analyses in both the EK and AK models $f = 0.50$ and $r = 0.00$ within one standard deviation, so these parameters were set at these values in subsequent runs. The upfield chemical shifts are 0.78, 0.74, 0.43, and 0.42 ppm, respectively, in the EK model and 0.90, 0.86, 0.50, and 0.48 ppm in the AK model.

The equilibrium constants in Table 4 found from the NMR chemical shift experiments agree well with those listed in Table 2 deduced from osmometry, better agreement occurring with the EK model. In the EK model with an assumed $\rho = 1$, a sedimentation equilibrium study[46] reports $K_E = 34$, only slightly greater than the other two values.

6. 1,N^6-Ethenoadenosine

In this fluorescent nucleoside an ethylene group joins the N1 and anilino N6 nitrogen of the adenine to form an additional fused five-membered ring with new hydrogens H10 and H11. Data are available for five protons: H2, H8, H10, and H11 on the nucleic base, and H1′ on the ribose.[47] The most general nonlinear least-squares analysis with simultaneous inclusion of all five protons indicates that within one standard deviation $f = 0.50$ and $r = 0.00$, so these parameters are set at these values in subsequent runs. The equilibrium parameters appear in Table 4 and the upfield shifts from monomer to molecule within a stack are 1.27, 0.71, 0.93, 0.79, and 0.65 ppm, respectively. With $\rho = 1.00$ assumed in the literature the average $K_E = 9.4 \pm 1.2$ calculated separately over each nucleus, and the upfield shifts are 11% greater than those above. Again, the ρK_E products of the two calculations are virtually identical. Results for the AK model also appear in Table 4 and the upfield shifts are 9% greater than those above. For common nuclei, upfield shifts for 1,N^6-ethenoadenosine are appreciably greater than those for N^6,N^6-dimethyladenosine, though the latter stacks about twice as strongly (Table 4).

The fact that the 1,N^6-ethenoadenosine equilibrium constants found by NMR are only half those found by osmometry has received comment,[47,48] and the discrepancy is only slightly reduced by the more thorough analyses performed here and reported in Tables 4 and 2. The fit by osmometry is the worst of the entries in Table 2. We have also reworked heat of dilution data[49] and used the protocol of section VI with ρ set to 0.8 to calculate $K_E = 20(11)$, which supports the greater osmometry result but without much conviction, owing to the large standard deviation. Possibly both osmometry and NMR results are correct; the two techniques might sense a different ensemble average as mentioned in the fourth paragraph of this article. However, there are other discrepancies of two or so and it is difficult to rule out experimental errors.

7. 5′-H(AMP)$^-$

From a careful study of 5′-AMP stacking at several pH values we have selected that at pH 5.2 (pD 5.6) where 90% of the molecules exist in a neutral base and monoprotonated, single negative phosphate charged form,[50] allowing comparison with results from the sedimentation equilibrium study performed in the same pH region. In the most general analysis of the proton NMR results we minimize the error in the H2, H8, and H1′ protons simultaneously and find that within one standard deviation $f = 0.50$ and $r = 0.00$ in both the EK and AK models. With these parameter set at these values we obtain the results listed in Table 4. Since ρ is nearly unity, the equilibrium constant K_E and the upfield shifts from monomer to molecule within a stack in the EK model of 0.84, 0.43, and 0.42 ppm, respectively, are virtually identical to those originally reported. Upfield shifts under the AK model are 30% greater.

The proton NMR 5′-H(AMP)$^-$ result obtained at pH 5.2 may be compared with that from sedimentation equilibrium performed in the same pH region with the results listed in Table 1. Under the EK model, $K_E = 3.4$ from NMR and 5.3 from sedimentation, and $\rho = 1.03$ from NMR and a much greater 1.92 from sedimentation. Except for a probably errant value for inosine in Table 2, the last value is the greatest encountered for ρ in this study. Thus the ρK_E product is 3.4 from NMR and 10.2 from sedimentation.

Under the AK model the τ values are both unity but $K_A = 5.3$ from NMR and a much greater 18 from sedimentation, again a factor of 3 or more for sedimentation. This difference from two careful studies is the most significant between two techniques to appear in this article. In the NMR study the highest concentration was 0.40 M; limiting the sedimentation data to 48 points to the same concentration yields $K_E = 5.07(6)$, $\rho = 2.12(6)$, and $B = 0.74(2)$, results which slightly worsen the difference in the $\rho K_E = 10.7$ product with that from proton NMR. Unlike most other examples 5′-H(AMP)$^-$ is charged, and in both experiments the ionic strength increased with concentration, ruling out an easy explanation on this basis for the difference. Care was taken in the sedimentation equilibrium study to accommodate the charged species, but this is difficult to do. Alternatively, the difference is a real feature of the two techniques measuring different aspects of self-association as mentioned in the fourth paragraph of this article.

8. Mg(ADP)$^-$ and Mg(ATP)$^{2-}$

Experiments on these two magnesium complexes were performed near pH 7 where there is Mg^{2+} and no proton on the phosphates and no interaction of the metal ion with the nucleic base. The charge neutralization provided by Mg^{2+} increases the stacking constants and more than half the complexes appear as trimers or higher stacks in the most concentrated solutions. Thus these complexes provide a test of the values of f and r in the nonlinear least-squares analysis performed simultaneously on three protons: H2, H8, and H1′. For both complexes $f = 0.50$ and $r = 0.00$ within one standard deviation, and as they are poorly specified in the most general analysis, these parameters are set to these values in subsequent calculations yielding the results in Table 4. With ρ taken as unity the original papers report for Mg(ADP)$^-$, $K_E = 6.4$,[51] and for Mg(ATP)$^{2-}$, $K_E = 4.0$.[42] These values are close to the ρK_E products in Table 4 of 5.6 and 3.5, respectively. The low ρ values in the EK model have as their counterpart low τ values in the AK model. Upfield shifts from monomer complex to complex in the interior of a stack are similar for the two nucleotides at 0.68, 0.37, and 0.34 ppm, respectively, in the EK model and at 0.80, 0.45, and 0.40 ppm in the AK model. These values are less than those reported in the literature under the EK model with $\rho = 1.00$.

Extensive NMR chemical shift stacking results from the isodesmic analysis of nucleosides and nucleotides and their metal ion complexes have received review.[52] The same order of decreasing stacking tendency prevails for each class of compound as for the nucleosides: adenosine > guanosine > inosine > cytidine > uridine, with the last two closely similar. For the derivatives of a single base the order of decreasing equilibrium constants as for adenosine is adenosine > Mg(ADP)$^-$ > Mg(ATP)$^{2-}$ > AMP^{2-} > ADP^{3-} > ATP^{4-}, consistent with increasing charge repulsion in stacks. Mg^{2+} binds only at the phosphate groups, and owing to a slightly enhanced stability, it has been suggested that the metal ion bridges phosphates from two ligands. However, the low ρ and τ values for the complexes (Table 4) indicates that the enhanced tendency to dimerization does not hinder significantly formation of longer stacks.

9. ϵ-ATP^{4-} and Mg(ϵ-ATP)$^{2-}$

With its negatively charged group the triphosphate derivative of 1,N^6-ethenoadenosine (ϵ-ATP) stacks more weakly than the nucleoside. Setting $f = 0.50$ and $r = 0.00$ we find by simultaneous analysis of the five protons, H2, H8, H10, and H11 on the nucleic base, and H1′ on the ribose, the results listed in Table 4 with upfield shifts in the EK model of 1.28, 0.66, 1.02, 0.87, and 0.57 ppm, respectively. In the original paper with assumed $\rho = 1$, for calculations over individual nuclei the average $K_E = 1.9 \pm 0.2$ and the upfield shifts are 27% greater than above.[53] Again, the ρK_E product is similar in the two calculations. In the AK model the upfield shifts are 20% greater than those above.

Particular attention is focused on the Mg^{2+} complex of ϵ-ATP (with Mg^{2+} only at the phosphate[54]) because stronger stacking results in concentrated solutions of greater than 50% of molecules in stacks of three or more molecules. A complete, simultaneous nonlinear least-squares analysis of the same five protons as above yields within one standard deviation $f = 0.50$ and $r = 0.00$. This result is important because, with those in the previous section, to an excellent approximation it suggests that these parameters may be set at these values. Doing so we obtain the results in Table 4 with upfield shifts from monomer to internally stacked molecule in the EK model of 1.20, 0.71, 0.98, 0.83, and 0.61 ppm, respectively. The original paper assumed $\rho = 1$ and calculated each nucleus separately to obtain the average $K_E = 7.6 \pm 0.5$ and upfield shifts 8% greater than above.[53] Again, the ρK_E product is virtually identical in the two calculations. In the AK model the upfield shifts are 17% greater than those listed above.

10. Cytidine and Uridine

These two pyrimidine nucleosides exhibit $L < 0.8$ in concentrated solutions implying about $^1/_4$ of the molelcules in a stack of three or more bases so we set $f = 0.50$ and $r = 0.00$ making the two treatments equivalent. For both nucleosides data are available for three nuclei, H5, H6, and H1′ on the ribose ring.[42] The three nuclei are considered simultaneously in a nonlinear least-squares fit to give the results in Table 4. For cytidine the upfield shifts from monomer to interior stack molecule are only 0.20, 0.11, and 0.10 ppm, respectively. For uridine they are even less at 0.11, 0.05, and 0.11 ppm, respectively. (All shifts possess a standard deviation of 0.01 ppm.) These small shifts make difficult accurate determination of the equilibrium constants. (For many years it was thought there were no shifts in pyrimidines due to stacking.) Since ρ is nearly unity the results agree well with the original paper where the three nuclei were processed individually. Both models yield the same apparent σ of 0.0013 ppm for cytidine and 0.0010 ppm for uridine but the AK model struggles with an unstable determinant in processing shifts for the weakly shifted stack molecules. Listed results

in Table 4 for both compounds conform to eq 35 relating the EK and AK models.

Both the EK and AK results of Table 4 are in the same range as those from sedimentation in Table 1 and osmometry in Table 2. We have for cytidine $1.1 < K_E < 1.6$ and $0.5 < \rho < 0.9$, and $3.5 < K_A < 6.1$ and $0.2 < \tau < 0.5$, and for uridine $0.7 < K_E < 1.1$ and $0.6 < \rho < 0.9$, and $2.2 < K_A < 4.1$ and $0.3 < \tau < 0.6$.

11. 1,10-Phenanthroline

Another case for which the maximum $K_E C_T = L > 2$, and hence more than half the molecules occur as higher than dimers, is that for *o*-phenanthroline in 25% methanol.[41] This reference suggests for protons under these conditions that $0.6 < f < 0.9$, and instead of being vertical, the stack is stepped such that a stacked molecule does not possess two equivalent nearest neighbors. Using the same data of 9 points to 0.4 M and allowing ρ to vary and fitting the four protons simultaneously, we find an excellent fit with an apparent σ of 0.01 ppm, $K_E = 6.4(1)$, and $\rho = 0.72$-(1), and for the H_α, H_β, H_γ, and H_δ protons, $f = 0.49$-(1), 0.39(1), 0.48(1), and 0.54(1), respectively. These values of $f \leq 0.5$ make it unnecessary to propose a stepped rather than vertical stack. By our analysis the upfield chemical shifts for the four protons are 0.9, 1.3, 1.8, and 2.1 ppm, respectively.

Coupled with those of Table 3, these results suggest that in a neutral molecule a typical high proton upfield shift from monomer to interior of a stack is about 2.0 ppm and that protons on stacked molecules may display lesser upfield shifts. Therefore, an upfield shift greater than 1 ppm for neutral molecules indicates that association proceeds beyond the dimer stage.

In water phenanthroline solubility limits the maximum $L \approx 0.9$ for which about 30% of the molecules are in stacks of three or more molecules at the highest concentration. For both the EK and AK models nonlinear least-squares fitting to the four protons simultaneously indicates that within one standard deviation $f = 0.50$ and $r = 0.00$. Therefore, these parameters are set at these values and for 12 points to 0.023 M[55] we obtain excellent fits with the results listed at the bottom of Table 4. In the EK model we find for the upfield shifts of the four protons: 0.62(4), 0.77(4), 1.32(7), and 1.71(9) ppm, respectively. In the original paper where it is assumed that $\rho = 1.00$, for calculations over individual nuclei, the average $K_E = 31 \pm 3$, with upfield shifts about 14% greater than above. The ρK_E product is virtually identical in the two calculations. In the AK model we obtain upfield shifts of 0.71(3), 0.87(3), 1.50-(4), and 1.94(4) ppm, respectively. Again, the comparison of the upfield shifts indicates that the two models yield different if similar solutions to the same set of data.

12. Daunorubicin

Daunorubicin (daunomycin) and doxorubicin (adriamycin) are closely related anthracycline antibiotics with an amino sugar, daunosamine. The structures differ in only one ring substituent: doxorubicin is 14-hydroxydaunorubicin, and this difference is without consequence for our purposes. The pair of compounds have been used interchangeably in investigations of their ability to form dimers and possibly stack. The acid−base and metal ion-binding properties of the compounds have been described.[56,57] A claim that only the ammonium deprotonated adriamycin dimerizes is based upon incorrect acidity constants and faulty interpretation.[58] The equilibrium constants for self-association are 1000 times greater than other compounds of this review and association has already occurred below the readily accessible concentration ranges of some of the techniques described above. NMR experiments were performed in solutions where association has already mainly taken place necessitating long and uncertain extrapolations to infinite dilution. Within the same paper NMR chemical shift and ultraviolet absorption intensity gave discordant dimerization constant values.[59] Optical spectroscopy provides the sensitivity to detect association in the 0.02 to 2 mM range. Absorption spectra provide lesser intensity changes than multipeaked circular dichroism spectra, which furnishes the most reliable tool.

As mentioned at the beginning of section IX, the formalism for the concentration dependence of NMR chemical shifts is identical to that for intensities in optical spectroscopy. The most dramatic changes in the circular dichroism spectrum with daunorubicin concentration changes occur at 350, 450, and 515 nm for which intensity data are available for 17 points from 0.02 to 1.56 mM, a range that covers most of the self-association.[60] We analyze this data by nonlinear least squares simultaneously at the three wavelengths. The most general analysis indicates that both $f = 0.50$ and $r = 0.00$ within one standard deviation. This result suggests that next nearest neighbor interactions are insignificant in optical spectroscopy. With the above parameters set at the indicated values we obtain in the EK model $K_E = 14.4(4)$ mM^{-1} and $\rho = 1.00(5)$, and in the AK model $K_A = 28(6)$ mM^{-1} and $\tau = 0.7(2)$.

As indicated at eq 43, with $\rho = 1.00$, the isodesmic and dimer models are formally indistinguishable; if only dimerization occurs, $K_2 = 7.2(2)$ mM^{-1}. This value agrees well with the original paper in which the data were reported for pH 7, 25 °C and 10 mM phosphate.[60] Increasing the ionic strength to 0.2 M increases the dimerization constant of the cation to about 10 mM^{-1}.[57] An identical dimerization constant was obtained indirectly from an analysis of Fe(III) binding to daunorubicin at pH 7.0 and 20 °C.[56] The partial stacking of daunorubicin results in the binding of a second ligand to a metal ion to be stronger than the first.[56]

Because the authors inexplicably and unnecessarily assumed that the dimer concentration is low compared to monomer in a region where dimer dominates, an equilibrium constant estimated from circular dichroism is grossly undervalued.[61] From the inset of their Figure 3, the midpoint occurs near 0.3 mM in a transformation that has not yet leveled off at the lowest concentration; eq 36a requires that the dimerization constant $K_2 > 5$ mM^{-1}, consistent with the above. As was also the case above, NMR data in this study were taken in a concentration range where dimerization had already occurred.

E. NMR Conclusions

There appears to be no correlation between the equilibrium constants for self-association, K_E and K_A, and the magnitude of the upfield shift from monomer to molecule within a stack. This point was mentioned in section IX.D.2 in comparing 6-methylpurine with purine. Adenosine stacks more than twice as strongly as $Mg(ADP)^-$, $Mg(ATP)^{2-}$, and 5′-H$(AMP)^-$ (Table 4), yet displays less than 70% of the upfield shift. The strongly self-stacking N^6,N^6-dimethyladenosine exhibits an average upfield shift. The three compounds derived from 1,N^6-ethenoadenosine give significantly greater upfield shifts than any other adenosine derivative but display ordinary K_E and K_A values in Table 4. Inosine with the smallest upfield shift of the purines in Table 4 exhibts an average stacking constant. Purine itself displays the greatest upfield shift coupled with the lowest stacking constant of any purine in Table 4. Other comparisons not appearing in this article support the generalization that no correlation exists between upfield shift magnitude and self-association equilibrium constant. Evidently, the requirements for strong stacking and high upfield shifts are not identical.

One of the features of this article is the focus on the extent to which next-nearest-neighbor interactions contribute to upfield shifts from monomer to molecules at the end of stack and to molecules within a vertical stack (sections IX.C and IX.D). Except for the early purine case of Table 3, we find $f = 0.50$ and $r = 0.00$ within one standard deviation, a result suggesting minimal next nearest neighbor interactions. Especially convincing are the cases of N^6,N^6-dimethyladenosine, $Mg(ADP)^-$, $Mg(ATP)^{2-}$, $Mg(\epsilon\text{-}ATP)^{2-}$, and 1,10-phenanthroline in 25% methanol where in the more concentrated solutions greater than 50% of the molecules reside in stacks of three or more molecules. These five examples exhibit substantially lower standard deviations than the analyses of the purines. Even for the purines assumption of $f = 0.50$ and $r = 0.00$ produces only marginally inferior fits. To the extent that results on the five compounds may be used as a test for a r value, it is a small number, perhaps about $r = 0.05$ or less ($f = 0.48$). For the most accurate work, this value might be assumed as a constant, but the difference with negligible next nearest neighbor interactions is so small that it does not seem worth the trouble. However, the value should be tested by more examples where there is a high proportion of molecules in stacks longer than dimer. A repeat of the purine and 6-methylpurine cases would be helpful. One might still wish to include a variable f in the nearest-neighbor treatment of section IX.B as it allows for stack features in addition to those related to r through eq 52.

Often upfield shifts calculated with the AK model are greater than those found from the EK model. The short extrapolation to zero concentration yields virtually identical monomer shifts in the two models; the difference arises in the longer extrapolation for the shift of interior stack molecules where only a small fraction of the molecules reside. This difference is disturbing because the chemical shift for a molecule in the interior of a stack should be a definite quantity independent of model. To the extent that the two models yield different upfield shifts, at least one of them must be incorrect. In at least one of the models the nonlinear least-squares minimization procedure leads to a slightly incorrect upfield shift and consequently slightly incorrect equilibrium constants.

X. Incorporation of a Second Molecule

To a distribution of molecule A of known stack lengths is added a small amount of molecule B. Molecule A is in excess and its distribution is undisturbed by addition of a small amount of B. (If one component is not present in excess a more elaborate treatment is necessary.[62]) Added molecule B may appear as a monomer, B, at the end of a stack, AB, or within a stack, ABA. We define the mole fraction of molecule B in these three environments as a, b, and c respectively; $a + b + c = 1$. There are two different ways to set up the interaction of B with a distribution of excess A.

First we consider addition of ends of species A_e to B according to

$$B + A_e \rightleftharpoons BA_e \qquad K_1 = b/a[A_e]$$

$$BA_e + A_e \rightleftharpoons A_eBA_e \qquad K_2 = c/b[A_e]$$

where $[A_e]$ is the molar concentration of ends of species A. Statistically one expects $K_1 = 4K_2$. For the weight fraction of A_e we have from section IX.B.1

$$[A_e]/C_T = 2\alpha + \lambda = 2(1 - \alpha L)$$

and hence $[A_e]$ is known for any concentration C_T. From the definitions it may be shown that the mole fractions of molecule B are described by

$$a = 1/D \qquad b = K_1[A_e]/D \qquad c = K_1K_2[A_e]^2/D$$

where the denominator $D = 1 + K_1[A_e] + K_1K_2[A_e]^2$.

The second approach is novel and considers substitution of molecule B for a molecule of A in its end, RA_e, and internal environments, RA_iR.

$$B + RA_e \rightleftharpoons RB_e + A \qquad K_e = b\alpha/a\lambda$$

$$B + RA_iR \rightleftharpoons RB_iR + A \qquad K_i = c\alpha/a\xi$$

If the second molecule B binds to or in a stack of A molecules with the same strength as another A molecule, then the values of the dimensionless equilibrium constants K_e and K_i are unity. Statistically we expect $K_i = K_e^2$. It may be shown that the mole fractions of molecule B are given by

$$a = \alpha/E \qquad b = \lambda K_e/E \qquad c = \xi K_i/E$$

where the denominator $E = \alpha + \lambda K_e + \xi K_i$.

The two formulations are related by the equations $K_e = K_1/K$ and $K_i = 4K_1K_2/K^2$, where K is an association constant for excess component A.

Both formulations yield expressions for mole fractions of component B in its three environments: free monomer, at the end of a single or stack of A molecules, and within a stack of A molecules. These mole fractions have been designated a, b, and c, respectively. An observable time-averaged property

such as the chemical shift of species *B* not present in excess is given by an expression analogous to eq 37

$$Q = aQ_a + bQ_b + cQ_c \qquad (53)$$

where Q_a, Q_b, and Q_c are the values of the property for monomer, molecule *B* at the end of a stack of A , and molecule *B* within a stack of A, respectively.

Through the equations for the mole fractions this equation expresses the observable Q in terms of parameters to be determined, K_1 and K_2 in the first formulation or K_e and K_i in the second. We may use either the EK or AK model for the distribution of A molecules present in excess.

Data have been published for incorporation into purine stacks of caffeine (1,3,7-trimethylxanthine) and three related indole derivatives: 5-hydroxyindole, tryptamine = 3-(2-aminoethyl)indole, and serotonin = 5-hydroxytryptamine, that are at low enough concentration of 7−8 mM so that they do not self-stack.[63] In the original paper equations related to the first approach above were manipulated to give a straight-line plot from which parameters were deduced from slope and intercept. The authors concluded that the three indole derivatives associate as strongly with stacked purine as does purine itself.

We have applied eq 53 in a nonlinear least-squares analysis simultaneously to five sets of protons in 5-hydroxyindole, three in tryptamine and caffeine, and four in serotonin. The precision of the data and the limited range of the purine concentration, to only 0.5 M in the last three cases making only about $^1/_3$ of the purine in stacks longer than dimers at the highest concentration, require that we assume the usual $f = 0.50$ and the statistical relation $K_i = K_e^2$. In the EK model for the purine distribution we used $K_E = 2.2$ and $\rho = 1.00$, the same values as in the original paper, and in the AK model the values atop Table 4. Our results for the dimensionless, substitution at stack end equilibrium constant K_e appear after each compound, first the value from the EK model and second from the AK model, 5-hydroxyindole, 0.98(5) and 1.05(5); tryptamine, 1.33(8) and 1.48(8); serotonin, 1.33(7) and 1.49(8); and caffeine, 1.40(7) and 1.54(5). Somewhat greater values are obtained in the AK model. The indoles self-associate more weakly than purine[63] and more weakly than they associate with purine. With three methyl groups, caffeine, on the other hand, self-associates (section VII) more strongly than it associates with purine. The largest upfield shifts of 1.55 ppm were found for the across six-membered ring 4- and 7-protons in 5-hydroxyindole within a purine stack. Such high values are consistent with indole insertion into a purine stack.

From the average K_e for each indole derivative and caffeine we may calculate from the statistical relation the following K_i values for substitution of a molecule for purine within a purine stack: 5-hydroxyindole, 1.03; tryptamine, 1.97; serotonin, 1.99; and caffine, 2.16. Thus while the first compound substitutes into a purine stack with the same facility as purine itself, we find that the last three compounds do so twice as strongly. This new result indicates that the 3-(2-aminoethyl)- substituent makes tryptamine and serotonin enter a purine stack with twice the strength of 5-hydroxyindole. This conclusion is consistent with the behavior of other alkyl substituted molecules: compare the two times greater stacking constant for 6-methylpurine over purine in Tables 2 and 4, and N^6,N^6-dimethyladenosine over (deoxy)-adenosine in the same two tables.

Purine to 0.90 M was added in excess to 0.010 M adenylyl(3′−5′)adenosine (ApA) and decreasing upfield shifts observed in four ApA base protons: H2-(5′), H2(3′), H8(3′), and H8(5′).[64] We have performed a nonlinear least-squares analysis from six concentrations read from a blowup of Figure 8 (from ref 64) and find $K_e = 1.4(1)$ and purine induced upfield shifts of 0.34, 0.33, 0.21, and 0.13 ppm, respectively. All four peaks from two different bases subscribe to the same equilibrium constant. The magnitude of the upfield shifts is much lower than shifts of the same nuclei in other adenine compounds of section IX.D.

Over the years at least 13 studies have estimated the degree of intramolecular stacking in ApA and the results tabulated.[65] Converting the results to 25 °C, we find that there are three ranges with seven studies reporting 19−35% stacking, five 51−68%, and one 85%, with no evident trend with experimental method and perhaps a tendency for later work to favor higher values. In nicotinamide adenine dinucleotide (NAD) a potentiometric study established 44% folding at 25 °C,[32] and we expect greater folding in ApA. A proton NMR study concluded that 87% of the 2:1 complexes of ADP and ATP and Al(III) are stacked.[66] In this study there was no evidence of intermolecular association of 2:1 complexes.

We conclude that about $^2/_3$ of ApA molecules are intramolecularly stacked and none intermolecularly. Added purine may interact with ApA in three ways. It may intercalate between adenine bases stacked intramolecularly. It may interact preferentially with unstacked ApA, each base acting independently. The resulting small upfield shifts are due to the offset of ApA unstacking. These two explanations account naturally for the identical equilibrium constant for purine association with each adenine base. Finally, if purine interacts externally with stacked ApA, it must do so equally with both bases.

XI. General Conclusions

In both the equal equilibrium constant (EK) and attenuated equilibrium constant (AK) models there is generally good agreement among the stacking equilibrium constants deduced for the same compound from very different experimental methods. Indeed, the good agreement may be interpreted as resounding reassurance for the whole enterprise of calculating stacking equilibrium constants from any of several physical methods. Where significant differences do occur, there has not been enough repetition of the experiments in different laboratories to justify reaching conclusions as to significance of the differences. The cases of 1,N^6-ethenoadenosine in section IX.D.6 and 5′-H(AMP)$^-$ in section IX.D.7 are offered as possible examples of real differences in constants between two techniques, suggesting conclusions such as those discussed in the fourth paragraph of this review.

Both the EK and AK models possess advantages and disadvantages. With its ever decreasing equilibrium constants the AK model allows for a greater improbability and entropy loss with a growing stack.[1] The EK model assumes the entropy changes remain constant regardless of stack length, thereby weighting more heavily the occurrence of long stacks.

On the other hand, comparison of the dimerization, K_2, and trimerization, K_3, constants in the simplest version of the two models shows $K_2 = K_3$ in the EK model and $K_2 = 1.5K_3$ in the AK model. However, we might expect $K_2 < K_3$ as there should be more entropy loss in dimer formation than in elongation of existing stacks because two monomers lose orientational freedom in the former and only one in an elongation.[6] Thus, on this point, the AK model is significantly more out of line with expectation. An adjustment may be made in both models by incorporating a new parameter $\rho = K_2/K_3$ in the EK model and $\tau = K_2/K_3$ in the AK model (compare eqs 12 and 19, see also section VIII). In this way the AK model may be described as "fixed".

These expectations are borne out by the analyses of this article. Tables 1, 2, and 4 show a persistent tendency for $\rho < 1$ in the EK model and its counterpart $\tau < 0.67$ in the AK model (see eq 35). The tendency is so strong that analyses that yield $\rho > 1$ and $\tau > 0.67$ are suspect. Accenting the discussion of section VIII, the NMR results of Table 4 show the lowest ρ and τ values for the most complicated molecules entailing the largest entropy loss on dimer formation. To now all literature NMR studies have assumed $\rho = 1.00$ and have not used the AK model. Comparison of the results of Table 1, 2, and 4 with those in the literature show a similar product ρK_E; values of $\rho < 1$ in this article are compensated by lower K_E values in the literature. Allowance for $\rho \neq 1.00$ demands solution of the cubic eq 14 instead of the quadratic eq 11 with $\rho = 1.00$. The AK model requires iterative solutions no matter the value of τ. Thus facile use of the quadratic equation accompanying $\rho = 1.00$ in the EK model might appear to be a casualty of this study. However, for weak interactions where the population of trimers and higher stacked species is low and ρ is difficult to determine reliably, it is certainly most convenient to employ the EK model with $\rho = 1.00$, requiring only solution of a quadratic equation to give an indication of the extent of stacking. This has been the course of almost all literature studies, and constants calculated on this basis remain a useful quantitative basis for comparison.

After numerous computer runs it is evident in those cases where the number of parameters to be determined combined with limited precision of the data challenges the nonlinear least-squares analyses that the equations of the EK model handle better with a more stable determinant than those of the AK model. Unfortunately, even though incorporation of the additional parameter τ "fixes" the AK model, it suffers from being the more difficult to use in practice. It also is more difficult to manipulate mathematically and, as the text shows, yields fewer relationships among the parameters.

XII. Acknowledgments

I thank Professors Claude Hélène, Helmut Sigel, Paul O. P. Ts'o, and Kensal E. van Holde, and Drs. Doris M. Cheng, Stephen R. Martin, Paul R. Mitchell (deceased), and Kurt H. Scheller for furnishing tables of data, often many years after procurement. I am grateful to Professor Carl Trindle of the University of Virginia not only for adapting the nonlinear least-squares program to a Macintosh computer, but especially for expanding the scope of the program so that it handles several independent and dependent variables.

XIII. References

(1) Garland, F.; Christian, S. D. *J. Phys. Chem.* **1975**, *79*, 1247−1252. (The second C_1 term, just before the brackets in eq 7, should be omitted.)
(2) Heyn, M. P.; Nicola, C. U.; Schwarz, G. *J. Phys. Chem.* **1977**, *81*, 1611−1617.
(3) Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1969.
(4) Martin, R. B. *FEBS Lett.* **1992**, *308*, 59−61.
(5) Heyn, M. P.; Bretz, R. *Biophys. Chem.* **1975**, *3*, 35−45.
(6) Sarolea-Mathot, L. *Trans. Faraday Soc.* **1953**, *49*, 8−20.
(7) Rossotti, F. J. C.; Rossotti, H. *J. Phys. Chem.* **1961**, *65*, 926−930; 930−934.
(8) Ghosh, A. K.; Mukerjee, P. *J. Am. Chem. Soc.* **1970**, *92*, 6408−6412.
(9) Adams, E. T.; Williams, J. W. *J. Am. Chem. Soc.* **1964**, *86*, 3454−3461.
(10) Kim, H.; Deonier, R. C.; Williams, J. W. *Chem. Rev.* **1977**, *77*, 659−690.
(11) Scatchard, G.; Hamer, W. J.; Wood, S. E. *J. Am. Chem. Soc.* **1938**, *60*, 3061−3070.
(12) Beckerdite, J. M.; Wan, C. C.; Adams, E. T. *Biophys. Chem.* **1980**, *12*, 199−214.
(13) Van Holde, K. E.; Rossetti, G. P. *Biochemistry* **1967**, *6*, 2189−2194.
(14) Rossetti, G. P.; Van Holde, K. E. *Biochem. Biophys. Res. Commun.* **1967**, *26*, 717−721.
(15) Van Holde, K. E.; Rossetti, G. P.; Dyson, R. D. *Ann. N. Y. Acad. Sci.* **1969**, *164*, 279−293.
(16) Adams, E. T. *Biochemistry* **1965**, *4*, 1646−1654.
(17) Adams, E. T.; Wan, P. J.; Crawford, E. F. *Methods Enzymol.* **1978**, *48*, 69−154.
(18) Schellman, J. A. *Compt. Rend. Trav. Lab. Carlsberg, Ser. Chim.* **1955**, *29*, 223−229.
(19) Ts'o, P. O. P.; Chan, S. I. *J. Am. Chem. Soc.* **1964**, *86*, 4176−4181.
(20) Solie, T. N.; Schellman, J. A. *J. Mol. Biol.* **1968**, *33*, 61−77.
(21) MacInnes, D. A.; Shedlovsky, T. *J. Am. Chem. Soc.* **1932**, *54*, 1429−1438.
(22) Katchalsky, A.; Eisenberg, H.; Lifson, S. *J. Am. Chem. Soc.* **1951**, *73*, 5889−5890. (Dimerization constants given for the other carboxylic acids could not be verified by our analysis.)
(23) MacDougall, F. H.; Blumer, D. R. *J. Am. Chem. Soc.* **1933**, *55*, 2236−2249.
(24) Davies, M.; Griffiths, D. M. L. *Z. Phys. Chem. Neue Folge* **1954**, *2*, 353−362.
(25) Sigel, H.; Martin, R. B. *Chem. Rev.* **1982**, *82*, 385−426.
(26) Martin, R. B. *J. Am. Chem. Soc.* **1967**, *89*, 2501−2502.
(27) Stoessor, P. R.; Gill, S. J. *J. Phys. Chem.* **1967**, *71*, 564−567.
(28) Gucker, F. T.; Pickard, H. B. *J. Am. Chem. Soc.* **1940**, *62*, 1464−1472.
(29) Gill, S. J.; Downing, M.; Sheats, G. F. *Biochemistry* **1967**, *6*, 272−276.
(30) Farquhar, E. L.; Downing, M.; Gill, S. J. *Biochemistry* **1968**, *7*, 1224−1225.
(31) Lonnberg, H.; Ylikoski, J.; Vesala, A. *J. Chem. Soc. Faraday Trans. 1* **1984**, *80*, 2439−2444.
(32) Sovago, I.; Martin, R. B. *FEBS Lett.* **1979**, *106*, 132−134. (The entropy of folding for NAD summarized in the first paragraph needs a negative sign as appears later in the text, $\Delta S = -11$ kcal mol^{-1} deg^{-1}.)
(33) Newcomb, L. F.; Gellman, S. H. *J. Am. Chem. Soc.* **1994**, *116*, 4993−4994.
(34) Guttman, D.; Higuchi, T. *J. Am. Pharm. Assoc.* **1957**, *46*, 4−10.
(35) Thakkar, A. L.; Tensmeyer, L. G.; Hermann, R. B.; Wilham, W. L. *J. Chem. Soc., Chem. Commun.* **1970**, 524−525.
(36) Ts'o, P. O. P.; Melvin, I. S.; Olson, A. C. *J. Am. Chem. Soc.* **1963**, *85*, 1289−1296.
(37) Gilligan, T. J.; Schwarz, G. *Biophys. Chem.* **1976**, *4*, 55−63.
(38) Mitchell, P. R.; Sigel, H. *Eur. J. Biochem.* **1978**, *88*, 149−154.

(39) Giessner-Prettre, C.; Pullman, B.; Borer, P. N.; Kan, L.; Ts'o, P. O. P. *Biopolymers* **1976**, *15*, 2277−2286.
(40) Cheng, D. M.; Kan, L. S.; Ts'o, P. O. P.; Giessner-Prettre, C.; Pullman, B. *J. Am. Chem. Soc.* **1980**, *102*, 525−534.
(41) Mitchell, P. R. *J. Chem. Soc. Dalton Trans.* **1980**, 1079−1086.
(42) Scheller, K. H.; Hofstetter, F.; Mitchell, P. R.; Prijs, B.; Sigel, H. *J. Am. Chem. Soc.* **1981**, *103*, 247−260.
(43) Kim, S.-H.; Martin, R. B. *Inorg. Chim. Acta* **1984**, *91*, 19−24.
(44) Chan, S. I.; Schweizer, M. P.; Ts'o, P. O. P.; Helmkamp, G. K. *J. Am. Chem. Soc.* **1964**, *86*, 4182−4188.
(45) Broom, A. D.; Schweizer, M. P.; Ts'o, P. O. P. *J. Am. Chem. Soc.* **1967**, *89*, 3612−3622.
(46) Bretz, R.; Lustig, A.; Schwarz, G. *Biophys. Chem.* **1974**, *1*, 237−241.
(47) Scheller, K. H.; Sigel, H. *J. Am. Chem. Soc.* **1983**, *105*, 3005−3014.
(48) Sigel, H. *Chimia* **1987**, *41*, 11−26.
(49) Sakurai, M.; Morimoto, S.; Inoue, Y. *J. Am. Chem. Soc.* **1980**, *102*, 5572−5574.
(50) Tribolet, R.; Sigel, H. *Biophys. Chem.* **1987**, *27*, 119−130.
(51) Scheller, K. H.; Sigel, H. *J. Am. Chem. Soc.* **1983**, *105*, 5891−5900.
(52) Yamauchi, O.; Odani, A.; Masuda, H.; Sigel, H. *Metal Ions Biol. Syst.* **1996**, *32*, 207−270.
(53) Scheller, K. H.; Sigel, H. *Eur. J. Biochem.* **1986**, *157*, 147−153.
(54) Sigel, H.; Scheller, K. H. *Eur. J. Biochem.* **1984**, *138*, 291−299.
(55) Tribolet, R.; Malini-Balakrishnan, R.; Sigel, H. *J. Chem. Soc. Dalton Trans.* **1985**, 2291−2303.
(56) Kiraly, R.; Martin, R. B. *Inorg. Chim. Acta* **1982**, *67*, 13−18.
(57) Martin, R. B. *Metal Ions Biol. Syst.* **1985**, *19*, 19−52.
(58) McLennan, I. J.; Lenkinski, R. E.; Yanuka, Y. *Can. J. Chem.* **1985**, *63*, 1233−1238.
(59) Chaires, J. B.; Dattagupta, N.; Crothers, D. M. *Biochemistry* **1982**, *21*, 3927−3932.
(60) Martin, S. R. *Biopolymers* **1980**, *19*, 713−721.
(61) Barthelemy-Clavey, V.; Maurizot, J.; Dimicoli, J.; Sicard, P. *FEBS Lett.* **1974**, *46*, 5−10.
(62) Weller, K.; Schutz, H.; Petri, I. *Biophys. Chem.* **1984**, *19*, 289−298.
(63) Dimicoli, J.; Helene, C. *J. Am. Chem. Soc.* **1973**, *95*, 1036−1044.
(64) Chan, S. I.; Nelson, J. H. *J. Am. Chem. Soc.* **1969**, *91*, 168−183.
(65) Frechet, D.; Ehrlich, R.; Remy, P. *Nucleic Acids Res.* **1979**, *7*, 1981−1987.
(66) Wang, X.; Nelson, D. J.; Trindle, C.; Martin, R. B. Manuscript in preparation, 1996.

CR960037V