Biophysical Chemistry 19 (1984) 289–298
Elsevier

BPC 00849

THERMODYNAMICAL MODEL OF INDEFINITE MIXED ASSOCIATION OF TWO COMPONENTS AND NMR DATA ANALYSIS FOR CAFFEINE-AMP INTERACTION *

K. WELLER, H. SCHÜTZ and I. PETRI

Akademie der Wissenschaften der DDR, Zentralinstitut für Mikrobiologie und experimentelle Therapie, DDR-69 Jena, G.D.R.

Received 18th August 1983
Revised manuscript received 2nd December 1983
Accepted 29th December 1983

Key words: Mixed association; Caffeine-AMP interaction; NMR; Parameter estimation; Matrix rank analysis

This paper describes the model used to estimate the parameters of caffeine-AMP interactions from corresponding ^{1}H-NMR measurements and some methods of data analysis by which the applicability of the model has been checked. The model of mixed association is applicable to a mixture of any two substances A and C which exhibit indefinite aggregates in both self-association and mixed association. In aggregates, only nearest neighbour interaction is assumed. The model is described by three equilibrium constants: K_{AA} and K_{CC} (for self-association of A, or C, respectively), and K_{AC} (for mixed association).

1. Introduction

In order to obtain some insight into the mechanism of interaction of caffeine with the adenine bases of DNA, ^{1}H-NMR studies of caffeine with adenosine monophosphate (AMP) have been performed [2]. The protons involved are shown in fig. 1. The final aim of these studies was the derivation of structural models.

In NMR spectroscopy of interacting systems with a fast exchange of molecules between different states, the chemical shifts of a substance are weight-averaged means of the chemical shifts of this substance in its pure states. The extent to which the chemical shifts are modified by interaction is described by so-called interaction shifts. Their values for any proton involved in NMR measurement reflect the mutual arrangement of molecules in the aggregates. Comparison with interaction shifts calculated in a semi-empirical manner on the basis of the ring-current model leads to structural models [2]. As a rule, for experimental reasons it is impossible to obtain NMR spectra of the pure states. Therefore, we need to know the relationship between the chemical shifts on the one hand and the interaction shifts and concentration fractions of appropriate pure states on the other. This will be described in section 2.

The concentration fractions of the states are unknown in experiments. The total concentrations

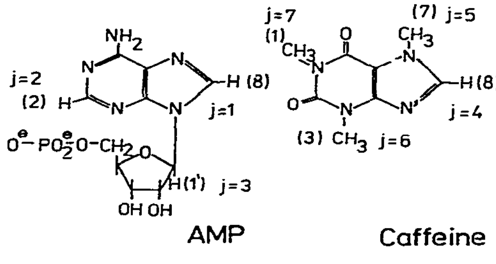

Fig. 1. The structural formulas of AMP and caffeine with the proton positions indicated that are involved in ^{1}H-NMR measurements. The assignment of number j to proton positions corresponds to the sequence in which the columns are arranged in tables 1 and 3.

---

* This paper and the following one [1] are theoretical counterparts to a paper [2] published in this journal.

0301-4622/84/$03.00 © 1984 Elsevier Science Publishers B.V.

of the substances are known primarily. Therefore, we need a thermodynamical model, with the aid of which we can calculate state concentrations as functions of total substance concentrations and equilibrium constants. An approximate formulation of the model of indefinite two-component mixed association is given by Hélène and Dimicoli [3]. Under the experimental conditions described in ref. 2, however, the assumption underlying this approximation is not fulfilled: We cannot assume that in all measured solutions, only a single AMP (or caffeine) molecule is present in each aggregate. Neither is the model of Von Tscharner and Schwarz [12] general to a sufficient extent to evaluate the present data. In section 3 we give exact formulas for our general model.

The whole model can be constructed from the formulas of sections 2 and 3. It describes the dependence of the measurable chemical shifts on the total concentrations and on the parameters, i.e., on the interaction shifts and the equilibrium constants. To estimate the parameters and the parameter errors, it is necessary to use a non-linear fitting programme and to formulate a subroutine for this programme on the basis of sections 2 and 3 as a model-specific part. Some computational aspects are also described in section 2. It is a matter of convenience in programming to subdivide the model-specific part of the programme into: (a) a subroutine related to the experimental method; and (b) a subroutine related to the interacting system studied.

Subroutine a defines system components (concentration fractions) and describes the dependence of the measured quantity (chemical shift) on these system components and on measurement parameters (interaction shifts). Subroutine b describes the dependence of the system components on independent variables (total concentrations) and on the system parameters (equilibrium constants). In this sense, section 2 deals with subroutine a and section 3 with subroutine b.

In section 4 some apparent properties of the model are discussed. Some justification for the choice of system components in section 2 comes from the matrix rank analysis of the data. Normally matrix rank analysis is applied to optical spectroscopy. The applicability of the matrix rank analysis and particularities of its application to NMR spectra are discussed in the appendix.

2. Connection between the chemical shifts and the system components

The goal of this section is to present the formulas necessary for the numerical procedure of the estimation of parameters, especially of the interactions shifts.

A method helpful in determining an adequate model from a set of spectroscopic data is matrix rank analysis. As described in the appendix, its application to a set of NMR spectra obtained during titration of caffeine-AMP mixed association and shown in table 1 yields a rank of 3 for both the chemical shifts of AMP protons alone and of caffeine protons alone, whereas for the total matrix the rank is 5. For self-association of

Table 1

Measured ^{1}HMR frequencies of solutions of Caffeine and AMP

Matrix δ_{ij} of the original test data of AMP-caffeine mixed association. The columns refer to the proton signals specified in fig. 1, and each row corresponds to a spectrum for the concentration $c^t_{A,i}$ indicated. $c^t_{C,i}$ was kept fixed at 0.13 M in each spectrum. The numbers of the third column are averaged values of the doublet frequencies. Measurements by Fritzsche et al. [2].

c^t_A (M)	H8A	H2A	H1'A	H8C	H7C	H3C	H1C
0				787.6	387.8	339.2	322.4
0.02	854.5	805.5	600.4	785.5	386.4	337.9	321.5
0.05	854.4	805.5	601.2	783.3	385.1	336.7	320.9
0.10	853.9	804.0	601.75	780.2	383.4	335.1	320.1
0.15	853.2	802.2	601.65	777.9	381.9	333.5	319.1
0.20	852.5	800.9	601.9	776.0	380.9	332.4	318.5

AMP as well as of caffeine, matrix rank analysis results in a rank of 2. In each case, the rank corresponds to the number of system components. Their identification is a matter of subsequent physical interpretation. As outlined in the appendix, these rank figures strongly suggest the introduction of the isolated states A and C and nearest neighbour interaction states AA, CC, AC as system components. These results further suggest for mixed association the following compositions of the chemical shifts of AMP protons (A protons) and of caffeine protons (C protons), containing three system components individually but five together. For A protons, i.e., for $j = 1$–3 (cf. fig. 1)

$$(\delta_{ji})_{\mathrm{A}} = \delta_{j.\mathrm{A}} + 2\Delta\delta_{j.\mathrm{AA}} c_{\mathrm{AA}.i}/c^{\mathrm{t}}_{\mathrm{A}.i} + \Delta\delta_{j.\mathrm{AC}} c_{\mathrm{AC}.i}/c^{\mathrm{t}}_{\mathrm{A}.i} \tag{1}$$

and for C protons, i.e., for $j = 4$–7

$$(\delta_{ji})_{\mathrm{C}} = \delta_{j.\mathrm{C}} + 2\Delta\delta_{j.\mathrm{CC}} c_{\mathrm{CC}.i}/c^{\mathrm{t}}_{\mathrm{C}.i} + \Delta\delta_{j.\mathrm{CA}} c_{\mathrm{CA}.i}/c^{\mathrm{t}}_{\mathrm{C}.i} \tag{2}$$

These expressions constitute the transposed matrix as compared to that of table 1, which has the same rank, of course. The subscript i enumerates the titration steps or the spectra. c_{AA}, c_{CC} and $c_{\mathrm{AC}} = c_{\mathrm{CA}}$ are the concentrations of AA, CC and AC neighbours, respectively. $c^{\mathrm{t}}_{\mathrm{A}}$ and $c^{\mathrm{t}}_{\mathrm{C}}$ are the total concentration of A and of C, respectively. $\delta_{j.\mathrm{A}}$ and $\delta_{j.\mathrm{C}}$ denote the chemical shifts of isolated A and C. The factor of 2 in eq. 1 describes the assumption that each of the two As in AA pairs with concentration c_{AA} contributes its interaction shift $\Delta\delta_{j,\mathrm{AA}}$. A similar explanation holds true for eq. 2.

For self-association, it is more convenient to refer to single molecules in aggregates rather than to molecule pairs as in eqs. 1 and 2 and to distinguish between inner and outer molecules in aggregates with concentrations c_{in} and c_{out}, respectively (refer to ref. 1). Inner molecules having two neighbours will experience twice as much nearest neighbour interaction chemical shift $\Delta\delta_{j.\mathrm{AA}}$ or $\Delta\delta_{j,\mathrm{CC}}$, respectively, compared with outer molecules having one neighbour, and the measured signal for self-association is

$$\delta_{ji} = \delta_{j,\mathrm{isol}} + \Delta\delta_j (c_{\mathrm{out},i}/c^{\mathrm{t}}_i + 2c_{\mathrm{in},i}/c^{\mathrm{t}}_i) \tag{3}$$

This expression contains only two independent system components, corresponding to a rank of 2. The expression in parentheses is equal to twice the relative concentration of molecule pairs in eq. 1 or 2, respectively.

For parameter estimation we used the criterion of least squares. Best values for the parameters are those, for which

$$\sum_{j=1}^{3} \sum_{i} (\delta_{ji} - (\delta_{ji})_{\mathrm{A}})^2 + \sum_{j=4}^{7} \sum_{i} (\delta_{ji} - (\delta_{ji})_{\mathrm{C}})^2 \tag{4}$$

is at a minimum. $(\delta_{ji})_{\mathrm{A}}$ and $(\delta_{ji})_{\mathrm{C}}$ have to be substituted by the right-hand sides of eqs. 1 and 2, respectively; δ_{ji} terms are the measured chemical shifts. For the calculation we used our general programme for non-linear regression on ALAU minicomputers [4] (H. Schütz, unpublished result), which runs reliably even with roughly guessed initial parameter values. Furthermore, we had at our disposal a general subroutine for spectroscopic measurements, MST [4] (measurement by spectroscopy of titrations), as subroutine a. The basic assumption is described by formula eq. A1 in the appendix. Measurements performed at any number of wavelengths may be evaluated for a system containing any number of components. By comparison of eq. A1 with eq. A3 it becomes clear that MST can also be used for the model described by eqs. 1 and 2. According to eqs. 1 and 2 the chemical shift of each proton is a homogeneous linear function of the measurement parameters ($\delta_{j.\mathrm{A}}$, $\Delta\delta_{j.\mathrm{AA}}$, $\Delta\delta_{j.\mathrm{AC}}$ for $j = 1$–3 or $\delta_{j.\mathrm{C}}$, $\Delta\delta_{j.\mathrm{CC}}$, $\Delta\delta_{j.\mathrm{CA}}$ for $j = 4$–7). We needed further a subroutine, which calculates $c_{\mathrm{AA}}/c^{\mathrm{t}}_{\mathrm{A}}$, $c_{\mathrm{AC}}/c^{\mathrm{t}}_{\mathrm{A}}$, $c_{\mathrm{CC}}/c^{\mathrm{t}}_{\mathrm{C}}$ and $c_{\mathrm{CA}}/c^{\mathrm{t}}_{\mathrm{C}}$ (the system components) and its derivatives by the thermodynamic parameters for each mixture of AMP and caffeine measured, as subroutine b. The formula basis of this subroutine is given in the next section, in which the thermodynamical model is treated.

3. Dependence of the system components on model parameters

The thermodynamical model of mixed association described below is a generalization of the type I SEK model in terms of ref. 5 or of the isodesmic model of self-association in ref. 6.

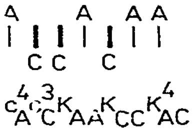

Fig. 2. The representation of a particular oligomer state in AC mixed association (as an example), with its statistical weight indicated.

3.1. Introduction of statistical weights

In terms of statistical thermodynamics, statistical weights of any oligomer state in mixed association have to be introduced. Because of their relative character the statistical weights may be identified with the respective concentrations. The statistical weight of any oligomer contains: (1) a factor c_A for each A; (2) a factor c_C for each C; (3) a factor K_{AA} for each AA neighbourhood; (4) a factor K_{CC} for each CC neighbourhood; (5) a factor K_{AC} for each AC neighbourhood.

Fig. 2 shows an example. c_A and c_C denote the concentration of isolated A and isolated C molecules, respectively, in solution. K_{AA}, K_{CC} and K_{AC} are the system parameters and denote the nearest neighbour equilibrium constants of association of A with A, C with C and A with C, respectively.

Distinction between K_{AC} and K_{CA} would force distinction to be made between c_{AC} and c_{CA} which contradicts the results of matrix rank analysis. More precisely speaking, neither eq. 1 nor eq. 2 may contain more than three system components. Undoubtedly, in eq. 1 c_{AC} could be replaced by $c_{AC} + c_{CA}$ and in eq. 2 c_{CA} by $c_{CA} + c_{AC}$. However, as seen from eqs. 13, 11 and its derivation, the partition function Z of the generalized model would depend symmetrically on K_{AC} and K_{CA} as would any logarithmic derivative and, consequently, any system component concentration. Therefore, even if $K_{AC} \neq K_{CA}$, their estimation from experimental data would not be unique because exchange of these parameters leads to an equally good representation.

The restriction to nearest neighbour interaction in the interaction model of mixed association is not a cogent assumption. However, as only nearest neighbour system components may be distinguished, the evaluability of long-range association constants by curve fitting would be doubtful.

3.2. The derivation of balance equations

Pure A aggregates of length i have the concentration

$$K_{AA}^{i-1} c_A^i \tag{5}$$

A aggregates of length i with C borders of any length (zero length inclusively) at both ends have altogether the concentration

$$K_{AA}^{i-1} c_A^i \left(1 + K_{AC} c_C + K_{AC} K_{CC} c_C^2 + \dots\right)^2 = K_{AA}^{i-1} c_A^i \left(1 + \frac{K_{AC} c_C}{1 - K_{CC} c_C}\right)^2 \tag{6}$$

These A aggregates, additionally with $i-1$ C insertions of any length (zero length inclusively), have altogether the concentration

$$\begin{aligned} c_{A,i} &= K_{AA}^{i-1} c_A^i \left(1 + \frac{K_{AC} c_C}{1 - K_{CC} c_C}\right)^2 \\ &\quad \times \left(1 + \frac{K_{AC}^2}{K_{AA}} c_C + \frac{K_{AC}^2}{K_{AA}} K_{CC} c_C^2 + \dots\right)^{i-1} \\ &= K_{AA}^{i-1} c_A^i \left(1 + \frac{K_{AC} c_C}{1 - K_{CC} c_C}\right)^2 \\ &\quad \times \left(1 + \frac{K_{AC}^2}{K_{AA}} c_C \frac{1}{1 - K_{CC} c_C}\right)^{i-1} \end{aligned} \tag{7}$$

Now, obviously, the total concentration of A molecules in solution is given by the balance equation

$$\begin{aligned} c_A^t = \sum_{i=1}^{\infty} i c_{A,i} &= \frac{\partial}{\partial \ln c_A} \sum_{i=1}^{\infty} K_{AA}^{i-1} c_A^i \left(1 + \frac{K_{AC} c_C}{1 - K_{CC} c_C}\right)^2 \\ &\quad \times \left(1 + \frac{K_{AC}^2}{K_{AA}} c_C \frac{1}{1 - K_{CC} c_C}\right)^{i-1} \end{aligned} \tag{8}$$

After some rearrangements the result is

$$\begin{aligned} G_A &= c_A^t \left(1 - c_A K_{AA} - \frac{c_A c_C K_{AC}^2}{1 - K_{CC} c_C}\right)^2 \\ &\quad - c_A \left(1 + \frac{K_{AC} c_C}{1 - K_{CC} c_C}\right)^2 = 0 \end{aligned} \tag{9}$$

which may be resolved for c_A. Corresponding relations exist for caffeine, leading to $G_C = 0$. Setting $K_{AC} = 0$ leads to the well known expression for isodesmic self-association [7]

$$c^t = c/(1 - Kc)^2 \tag{10}$$

The algorithm for obtaining the solution (c_A, c_C) of the system of implicit equations $G_A = 0$ and $G_C = 0$ for given values of K_{AA}, K_{CC} and K_{AC} is analogous to that described in the appendix of ref. 1.

3.3. *The concentrations of system components*

With the solution (c_A, c_C) being known, the relative system component concentrations may be calculated. They will now be derived using the partition function Z of the system.

Z is obtained by adding the statistical weights of all the states possible to a mixed oligomer. Those states containing at least one A molecule are comprised by

$$\sum_{i=1}^{\infty} c_{A,i} = \sum_{i=1}^{\infty} K_{AA}^{i-1} c_A^i \left(1 + \frac{K_{AC}c_C}{1 - K_{CC}c_C}\right)^2 \times \left(1 + \frac{K_{AC}^2 c_C}{K_{AA}(1 - K_{CC}c_C)}\right)^{i-1} \tag{11}$$

The statistical weights of pure C aggregates

$$\sum_{i=1}^{\infty} K_{CC}^{i-1} c_C^i = \frac{c_C}{1 - K_{CC}c_C} \tag{12}$$

must be supplied to obtain

$$Z = \sum_{i=1}^{\infty} c_{A,i} + \frac{c_C}{1 - K_{CC}c_C} \tag{13}$$

After some transformations the result is

$$Z = \frac{c_A + c_C - c_A c_C(K_{AA} - 2K_{AC} + K_{CC})}{1 - (c_A K_{AA} + c_C K_{CC}) + c_A c_C(K_{AA}K_{CC} - K_{AC}^2)} \tag{14}$$

As expected, Z is symmetrical in A and C.

By the rules of partition function formalism, the average number of AA pairs per aggregate is given by $\partial \ln Z/\partial \ln K_{AA}$ and the average number of As per aggregate by $\partial \ln Z/\partial \ln c_A$.

Thus, the relative concentration c_{AA}/c_A^t in eq. 1 is given by

$$\frac{c_{AA}}{c_A^t} = \frac{\partial \ln Z}{\partial \ln K_{AA}} / \frac{\partial \ln Z}{\partial \ln c_A} = K_{AA}c_A \tag{15}$$

provided interaction between oligomers in solution is neglected. The remaining relative concentrations are:

$$\frac{c_{AC}}{c_A^t} = \frac{\partial \ln Z}{\partial \ln K_{AC}} / \frac{\partial \ln Z}{\partial \ln c_A} = \frac{2c_C K_{AC}(1 + c_A(K_{AC} - K_{AA}))}{(1 + c_C(K_{AC} - K_{CC}))} \tag{16}$$

$$\frac{c_{CC}}{c_C^t} = \frac{\partial \ln Z}{\partial \ln K_{CC}} / \frac{\partial \ln Z}{\partial \ln c_C} = K_{CC}c_C \tag{17}$$

$$\frac{c_{AC}}{c_C^t} = \frac{\partial \ln Z}{\partial \ln K_{AC}} / \frac{\partial \ln Z}{\partial \ln c_C} = \frac{2c_A K_{AC}(1 + c_C(K_{AC} - K_{CC}))}{(1 + c_A(K_{AC} - K_{AA}))} \tag{18}$$

An independent combinatorial reasoning yielded the same results. For self-association, the number of oligomer ends is twice the number of oligomers having ends. Therefore,

$$c_{\text{out}} = 2\sum_{i=2}^{\infty} K^{i-1}c^i = \frac{2Kc^2}{1 - Kc} \tag{19}$$

$$c_{\text{in}} = c^t - c - \frac{2Kc^2}{1 - Kc} = \frac{K^2c^3}{(1 - Kc)^2} \tag{20}$$

By inserting eqs. 15–18 into eqs. 1 and 2 (or for self-association eqs. 10, 19 and 20 into eq. 3), the connection between NMR signals and parameters is established. It is linear for chemical shifts $\delta_{j,A}$ and $\delta_{j,C}$ and interaction shifts $\Delta\delta_{j,AA}$, $\Delta\delta_{j,AC}$, $\Delta\delta_{j,CC}$ and $\Delta\delta_{j,CA}$, but non-linear for K_{AA}, K_{CC} and K_{AC}. By virtue of the balance equations, c_A as well as c_C are functions of each of the parameters K_{AA}, K_{CC} and K_{AC}. When forming the derivatives of eqs. 15–18 by these parameters, one has to pay attention to this fact. The derivatives are needed in our general programme for non-linear regression because it is based on the GAUSS-NEWTON algorithm.

4. Discussion of the model behaviour

Parameter estimation was done on the basis of data displayed in fig. 1a of ref. 2. K_{AA}, $\Delta\delta_{j,AA}$, K_{CC} and $\Delta\delta_{j,CC}$ were obtained from self-association investigations of AMP and caffeine, respectively (table of ref. 2). After some gauging procedure the corresponding values of $\delta_{j,A}$ and $\delta_{j,C}$ were obtained from the same self-association results. All these values were kept fixed. The minimum of eq. 4 was sought with respect to K_{AC} and all $\Delta\delta_{j,AC}$ and $\Delta\delta_{j,CA}$. The results are shown in table 2 (already given in ref. 2). Fig. 1a of ref. 2 shows the fitted curves.

Once the system parameter values are known, the curves of system component concentrations vs. c_A^t, with c_C^t kept fixed at 0.13 M, may be constructed. Fig. 3 shows the corresponding plot. (The concentration range of c_A^t is the same as in fig. 1a of ref. 2.)

Fig. 3 shows the behaviour of the caffeine-AMP system at a fixed caffeine concentration and increasing concentrations of AMP. Initially, there is only the self-association of caffeine. We recognize the finite initial values of free caffeine concentrations and of CC neighbourhood frequencies. On addition of AMP, AC neighbourhoods are formed increasingly and, therefore, both c_C and c_{CC} decrease and c_{AC} increases, absolutely as well as relatively to caffeine (the total caffeine c_C^t concentration being constant). Initially, c_A and c_A^t are proportional to one another, as are c_{AC} and $c_A c_C$ as well as c_{AA} and c_A^2. For this reason, c_{AC}/c_A^t has a finite value at $c_A^t = 0$, but c_{AA}/c_A^t begins from a value of 0. Similarly, it is obvious that c_{AC}/c_A^t is steadily decreasing (the concentration of free caffeine c_C becoming smaller) and that c_{AA}/c_A^t is increasing.

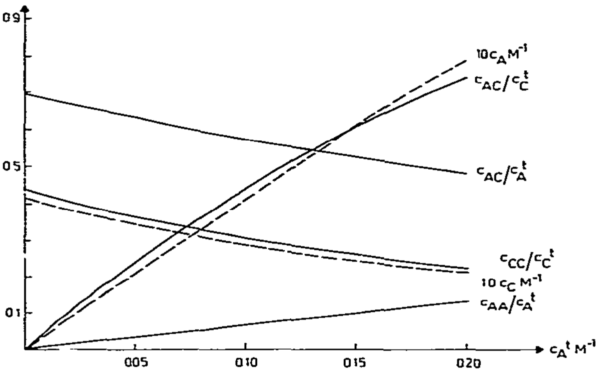

Fig. 3. The curves of the relative system component concentrations (eqs. 15–18) vs. c_A^t for the system parameter values estimated. $K_{AA} = 1.67\ M^{-1}$, $K_{CC} = 10.6\ M^{-1}$, $K_{AC} = 7.25\ M^{-1}$, with c_C^t kept fixed at 0.13 M (refer to table 2). c_{AA}/c_A^t, c_{AC}/c_A^t, c_{CC}/c_C^t, c_{AC}/c_C^t, $10c_A$, $10c_C$.

The model of mixed association applies to any interacting system consisting of two substances forming indefinite aggregates in both self-association and mixed association. In aggregates, nearest neighbour interaction, exclusively, is assumed and in mixed aggregates, any sequence of molecules of both kinds is allowed.

Table 2

Best estimates of parameter values and parameter errors (5% for $\Delta\delta_{AA}$, $\Delta\delta_{CC}$, $\leq 10\%$ for $\Delta\delta_{AC}$, $\Delta\delta_{CA}$)

Parameter values estimated of AMP-caffeine mixed association, the parameter error values being indicated. (Specifications of the A protons, $j = 1$–3, and of the C protons, $j = 4$–7, are indicated.)

	Chemical shifts and interaction shifts of mixed association caffeine-AMP (ppm)							Equilibrium constants (M^{-1})
	AMP			Caffeine				
	H8A	H2A	H1'A	H8C	H7C	H3C	H1C	
δ_A, δ_C	8.642	8.296	6.163	7.914	3.983	3.564	3.385	$K_{AA} = 1.67 \pm 0.17$
$\Delta\delta_{AA}$, $\Delta\delta_{CC}$	−0.204	−0.474	−0.144	−0.014	−0.119	−0.198	−0.181	$K_{CC} = 10.6 \pm 1.0$
$\Delta\delta_{AC}$, $\Delta\delta_{CA}$	−0.143	−0.347	−0.234	−0.183	−0.161	−0.204	−0.157	$K_{AC} = 7.25 \pm 1.19$

Appendix (by H. Schütz)

A1. On the applicability of matrix rank analysis to NMR spectra of interacting systems

In optical spectroscopy, the application of matrix rank analysis to the determination of the number of linearly independent spectroscopic components is not uncommon [8–11].

The absorbances are, apart from the errors, homogeneous linear functions of the extinction coefficients as well as of the concentrations

$$A_{ji} = \sum_{k=1}^{q} \epsilon_{jk} c_{ki} d_i \qquad i = 1,\ldots,n \qquad j = 1,\ldots,m \tag{A1}$$

where i is the index of the titration step, i.e., of the individual spectrum, j that of wavelength, and k that of system component. ϵ, c and d, as usual, denote extinction coefficients, concentrations, and path lengths, respectively.

The number of system components q is the unknown to be determined.

Eq. A1 can also be expressed in a matrix notation

$$\boldsymbol{A} = \boldsymbol{ECD} \tag{A2}$$

As $\boldsymbol{D}$ is a non-singular diagonal matrix (of the path lengths), according to a theorem of matrix calculus the rank of $\boldsymbol{A}$ is less than or equal to the minimum of the ranks of $\boldsymbol{E} = ((\epsilon_{jk}))$ and $\boldsymbol{C} = ((c_{ki}))$. For i kept fixed, eq. A1 constitutes a set of m linear equations for the unknowns c_k. As is well known, the number of c_k terms which can be determined is equal to the number of independent equations, and this in turn is equal to the rank of the matrix of the coefficients ϵ_{jk}. Of course, the number of system components does not depend on the titration step. Provided both the number of wavelengths, m, as well as the number of mixtures spectroscopically investigated, n (i.e., the number of different spectra), are large enough (in comparison with the expected value of q), and if the range of varying component concentrations among the mixtures is sufficiently extended, the rank of $\boldsymbol{A}$ may be identified with the number q of spectroscopically discernible system components.

If there were no errors, the rank of $\boldsymbol{A}$ would be easily accessible by the Gaussian algorithm (which physically means forming sets of difference spectra). In reality, however, careful analysis of the experimental errors is necessary to distinguish weakly pronounced components from noise.

For different reasons, within a set of spectra in our case an experimental error of 0.2–0.3 Hz must be assumed. This is the order of magnitude to which the absolutes of all the elements of the matrix $\boldsymbol{A}$ must diminish after a certain number of reduction steps if this reduced matrix is to have a rank of zero. This number of reduction steps equals the rank of the original matrix $\boldsymbol{A}$.

In the case of NMR spectroscopy, applied to interacting systems with a fast exchange of molecular states, two peculiarities have to be taken into account.

(1) For the chemical shifts δ_{ji} an expression of the type in eqs. 1 and 2 but not of the type in eq. A1 will be valid:

$$\delta_{ji} = \delta_{j,\mathrm{isol}} + \sum_{k=2}^{9} \Delta\delta_{jk} c_{ki}/c_i^{\mathrm{t}} \tag{A3}$$

j counts the different protons, k the interaction states, and i the spectra. $\delta_{j,\mathrm{isol}}$ refers to the isolated states of A or C, $c_i^{\mathrm{t}} = c_{\mathrm{A}i}^{\mathrm{t}}$ for A protons and $c_i^{\mathrm{t}} = c_{\mathrm{C}i}^{\mathrm{t}}$ for C protons.

The products $\delta_{ji} c_i^{\mathrm{t}}$, however, obviously show a structure according to eq. A1 or A2, as we may identify

$$\begin{aligned}
\boldsymbol{A}_{ji} &= \delta_{ji} c_i^{\mathrm{t}} && \\
\boldsymbol{E}_{jk} &= \delta_{j,\mathrm{isol}} && \text{for } k = 1 \\
&= \Delta\delta_{jk} && \text{for } k > 1 \\
\boldsymbol{C}_{ki} &= c_i^{\mathrm{t}} && \text{for } k = 1 \\
&= c_{ki} && \text{for } k > 1 \\
\boldsymbol{D}_{il} &= 1 && \text{for } i = l \\
&= 0 && \text{for } i \neq 1
\end{aligned} \tag{A4}$$

Then, from analogous considerations to those in eq. A2, the rank of $\boldsymbol{A}$ may be identified with the number q of states (comprising interaction states and isolated state) discernible by NMR spectroscopy.

The multiplication of all the elements of a column (i.e., of a spectrum) by a common factor (c_i^{t}) does not influence the rank of the matrix. Thus, the factor c_i^{t} may be omitted as long as the protons of one substance are analyzed alone. (On the other hand, factors may be used to amplify certain spectra.) If, however, the proton signals of different molecules (A and C in our case) have been combined to form a spectrum, $j = 1, \ldots, 7$, attention must be paid to the fact that the chemical shifts of the protons have to be multiplied by the respective total concentrations ($c_{\mathrm{A},i}^{\mathrm{t}}$ for A protons and $c_{\mathrm{C},i}^{\mathrm{t}}$ for C protons).

(2) In contrast to optical spectroscopy, in NMR it is not possible to measure at as many wavelengths as desired. The number of protons which show an interaction shift is very limited: three A protons and four C protons, in our case. From the chemical shifts of A protons only a rank not greater than 3, from those of C protons only a rank not greater than 4, and from those of A and C protons together a rank not greater than 7 may be evaluated.

It is not intended to discuss here the criteria which may be used to decide whether a matrix A, after having applied the Gaussian algorithm for a certain number of times, still has a rank significantly different from zero. At any rate, it will be a statistical criterion because of the experimental errors. Therefore, in cases of a limited number of protons that may be investigated as well of a limited variability in concentrations, methods which allow the spectroscopic components present to become more pronounced are especially valuable.

Such a possibility arises if there are components with previously known basic spectra which are known to be present in the mixture. Then, one can add these basic spectra to the set of spectra measured and apply matrix rank analysis to the whole set. If, moreover, one basic spectrum is amplified by a sufficiently large factor, the matrix rank analysis will begin by forming the differences between this basic spectrum (containing the pivot element) and all the other spectra. Obviously, these differences will be larger than those between any one of the mixture spectra and the others. This renders the distinction of system components from noise more significant. The effect of addition of basic spectra is demonstrated for the present NMR spectra of mixed association of AMP and caffeine.

Owing to the results of both AA and CC self-association (table 1 of ref. 2), in the concentration range indicated by table 1, isolated molecules as well as self-associated molecules of both A and C must be present in non-negligible fractions. Thus, we are justified in using the chemical shifts of the isolated molecules and the interaction shifts of self-association of both A and C as additional spectra in the matrix rank analysis. This is done in table 3 which corresponds to the transposed matrix of A in eq. A2. Note that this matrix has been derived from table 1 by multiplying the chemical shifts of the A protons by $c_{\mathrm{A},i}^{\mathrm{t}}/c_{\mathrm{C},i}^{\mathrm{t}}$.

A1.1. Analysis of A protons alone

Without the first row, we take columns 1–3 of table 1 as the spectra to be investigated. As a sixth row we add the 100-fold of the AA interaction shifts as shown in the seventh row of table 3. Then, we can be sure that this spectrum will be elected by the Gaussian reduction algorithm because it contains the pivot element. As the seventh and last spectrum we add the chemical shifts of the isolated A. However, this spectrum has been shifted by 0.6 Hz with respect to that given in table 1 of ref. 2 in order to assimilate it to the corresponding spectrum in AA self-association. It is shown in the ninth row of table 3.

To give an idea of the influence of the basic spectra added on the statements of the rank analysis, we look at the pivot elements before the third reduction step in the Gaussian algorithm for the matrices comprising columns 1–3 with rows 2–6, 2–7, and 2–7 and 9, respectively, in table 3. They are 0.9, 2.17 and 7.0, respectively.

Unambiguously – in comparison with the error assumed – the A spectra of the mixed association show the full rank of 3. The significance of this statement, however, is clearly enhanced by addition of the basic spectra.

As a result, besides the chemical shifts of the isolated A molecules and the AA interaction shifts one additional basic spectrum is needed for the description of the chemical shifts of the A protons alone in mixed association. This must be caused by the presence of C molecules in the mixtures.

Table 3

Matrix of chemical shifts, prepared for the common matrix rank analysis of AMP protons and caffeine protons

Matrix δ_{ij} with each row multiplied by $c^t_{A,t}/c^t_{C,t}$ for $j = 1$–3 and supplemented by all the basic spectra of self-association both of A and of C.

0	0	0	787.6	387.8	339.2	322.4
131.46	123.92	92.37	785.5	386.4	337.9	321.5
328.62	309.81	231.23	783.3	385.1	336.7	320.9
656.85	618.46	462.88	780.2	383.4	335.1	320.1
984.46	925.62	694.21	777.9	381.9	333.5	319.1
1311.54	1232.15	926.00	776.0	380.9	332.4	318.5
2035	4741	1444	0	0	0	0
0	0	0	411.6	1194.4	1976.7	1812.7
864.80	830.25	616.87	0	0	0	0
0	0	0	791.06	397.94	356.22	338.14

A1.2. Analysis of C protons alone

Analogously to section A1.1, we take columns 4–7 of table 1 and add the 100-fold of the CC interaction shifts and the gauged basic spectrum of isolated C, as displayed in the eight and tenth rows of table 3.

Now, we again compare the pivot elements before the third reduction steps of the matrices with the first six, the first seven, and all eight spectra, respectively, which are 0.30, 2.50 and 2.50, respectively. Before the fourth reduction step we find values of −0.22, −0.16 and −0.16. These results show that by addition of the CC interaction shifts it becomes clear that the matrix has a rank of 3. The spectra of the mixed association alone could ascertain only the rank of 2 because the third and the absolute value of the fourth pivot element (0.30, −0.22) are of the order of magnitude of experimental error. The addition of the second basic spectrum no longer enhanced the significance.

As a result, again, one and only one additional basic spectrum is needed (besides the basic spectra of CC self-association) for description of the chemical shifts of the C protons in mixed association. This must be caused by the presence of A molecules in solution.

A1.3. Joint analysis of A and C protons

Now, the whole content of table 3 is used for matrix rank analysis.

We again compare matrices with the first six spectra (mixed association alone), the first eight spectra (interaction shifts of self-association included), and of all ten spectra. For these three matrices, the pivot elements before the fourth reduction step are −0.38, 717 and 717, before the fifth reduction step −0.16, −0.89 and 6.24, and before the sixth reduction step −0.01, −0.12 and −0.16.

Thus, the rank of the six-row matrix is perhaps 4, the rank of the eight-row matrix is probably 5, but the rank of the full matrix is 5 and only 5. (Note that for parameter estimation of mixed association analogously the knowledge obtained from self-association is also included.)

As a result, from sections A1.1 and A1.2, we know that in mixtures one additional basic spectrum for A protons (caused by C) and one additional basic spectrum for C protons (caused by A) must be assumed. But for the chemical shifts of all the protons together, besides the four spectroscopic species known from self-associations (isolated A, AA neighbourhoods in aggregates, isolated C, CC neighbourhoods in aggregates) only one additional species corresponding to one additional basic spectrum must be assumed. The simplest and most satisfying assumption is the following. The fifth spectroscopic species is an AC neighbourhood. The corresponding basic spectrum over all the protons is composed of the interaction shifts of the A protons caused by one neighbouring C molecule and the interaction shifts of C protons caused by one neighbouring A molecule.

(The A part only and the C part only, respectively, of this basic spectrum are the basic spectra demanded in sections A1.1 and A1.2, respectively.) This is the assumption used in the model part of this paper.

References

1 K. Weller, H. Schütz and I. Petri, Biophys. Chem. 19 (1984) 289.
2 H. Fritzsche, I. Petri, H. Schütz, K. Weller, P. Sedmeřa and H. Lang, *Biophys. Chem. 11 (1980) 109.*
3 C. Hélène and J.-L. Dimicoli, J. Am. Chem. Soc. 95 (1973) 1036.
4 Fortran routines of all mentioned programmes exist.
5 J.M. Beckerdite, C.C. Wan and E.T. Adams, Biophys. Chem. 12 (1980) 199.
6 D. Pörschke and F. Eggers, Eur. J. Biochem. 26 (1972) 490.
7 L.-H. Tang, D.R. Powell, B.M. Escott and E.T. Adams, Biophys. Chem. 7 (1977) 121.
8 R.M. Wallace, J. Phys. Chem. 64 (1960) 899.
9 R.M. Wallace and S.M. Katz, J. Phys. Chem. 68 (1964) 3890.
10 D.M. Mc Mullen, S.R. Jaskunas and L. Tinoco, Jr, Biopolymers 6 (1967) 589.
11 I. Petri, W. Förster and G. Löber, Stud. Biophys. 45 (1974) 61.
12 V. von Tscharner and G. Schwarz, Biophys. Struct. Mech. 5 (1979) 75.