ASSOCIATION OF INOSINE AND CYTIDINE

M. E. MAGAR, R. F. STEINER, *and* R. KOLINSKI

From the Laboratory of Physical Biochemistry, Naval Medical Research Institute, Bethesda, Maryland 20014. Dr. Magar's present address is the Center for Theoretical Studies, University of Miami, Coral Gables, Florida 33124. Dr. Steiner's present address is the Department of Chemistry, University of Maryland Baltimore County, Baltimore, Maryland 21201.

ABSTRACT A rigorous theory for the analysis of an associating system involving two different monomer units, which may also undergo self-association, has been developed. In the present paper the application of the theory to actual data is demonstrated for the first time. The model system chosen is the interaction of inosine with cytidine, both of which self-associate strongly. Application of the theory indicates that inosine associates with cytidine and that the association constant is of a similar order of magnitude to that for the self-association of the monomers.

INTRODUCTION

The theory of mixed associating systems has been developed by Steiner (1968) and by Adams et al. (1969). In this manuscript we report the application of the former theory to the heterogeneous association of a mixture of inosine and cytidine with the objective of determining the equilibrium constant, or constants, of this associating system.

A primary purpose of the present paper is to demonstrate the applicability of this method of analysis to an actual associating system. The inosine-cytidine system is a good model system for this purpose since many of the complications dealt with by the theory are present, including self-association of both monomer species, as well as the lack of restrictions on the composition of the mixed complexes theoretically possible.

MATERIAL AND METHODS

Inosine and cytidine were purchased from Calbiochem, Los Angeles, Calif. All measurements were made with a Hewlett-Packard 302B vapor pressure osmometer (Hewlett-Packard Co., Palo Alto, Calif.). The output of this instrument was fed into a Keithley 150A microvolt ammeter (Keithley Instruments, Inc., Cleveland, Ohio) and the signal from the Keithley was fed into a voltage divider, thereby attenuating the output voltage of the amplifier by $\frac{1}{100}$ of its initial value. This signal was then recorded on a Hewlett-Packard 17504A recorder. The vapor pressure osmometer was calibrated with a National Bureau of Standards sucrose sample. The measurements were done at 50°C to take advantage of the greater solubility of

inosine at this temperature. Solutions were kept in a 50°C water bath before being transferred into the syringes of the osmometers. All least-square computations were made using the General Electric Time-Sharing Service (Bethesda, Md.).

THEORY, COMPUTATIONS, AND RESULTS

Two basic equations are involved in the analysis of a two-solute system. The first (Steiner, 1968) may be written as

$$\Phi(m, \beta) \equiv \ln X_A + \beta \ln X_B = (1 + \beta) \int_0^m \frac{(\alpha_n^{-1} - 1)\, dm}{m} + \ln \frac{1}{1 + \beta} + \beta \ln \frac{\beta}{1 + \beta}. \quad (1)$$

In this equation X_A is the mole fraction of monomer units of species A, X_B is the mole fraction of monomer units of species B. The constant β may be defined by the equation $m_{B,t} = \beta m_{A,t}$, where $m_{B,t}$ and $m_{A,t}$ are the total monomer concentrations of all B and A units, respectively, whether free or complexed; $\alpha_n = M_n/M_0 =$ number-average degree of association; and $m = m_t/\alpha_n$, where $m_t = m_{A,t} + m_{B,t}$. In the following material inosine will be labeled as A and cytidine as B. Equation 1 is used to compute the function $\Phi(m, \beta)$ for various values of m and β and use of this function will then yield the values of X_A and X_B which are subsequently used to evaluate the equilibrium constants using our second basic equation (equation 4, below).

To evaluate the function $\Phi(m, \beta)$ at a given β value, one makes serial dilutions of a mixture with that value of β. The number-average molecular weight for each one of these dilutions is measured. These values then yield α_n as a function of m. With these values of α_n and m one is then able to compute the integral on the right-hand side of equation 1 at various values of m. Once the integral is determined at various m, one can use the entire right-hand side of equation 1 to compute $\Phi(m, \beta)$ at each given value of β for a series of values of m, the end result being a tabulated function which for a given β, say β_1, will look like this: $\Phi(m_1, \beta_1)$, $\Phi(m_2, \beta_1)$, $\Phi(m_3, \beta_1)$, $\cdots$, etc. The same procedure is then followed for all other values of β.

In actual practice we plotted raw data determined as α_n and m for any given β and drew a smooth curve through the data. At suitable values of m the corresponding values of α_n were read off the smooth curve and these values were then used to perform the integration. This interpolation procedure was necessary because, in tabulating the function $\Phi(m, \beta)$, we essentially need the *same* values of m for the various values of β. An example of a smoothed plot of α_n vs. m at $\beta = 2.6$ is shown in Fig. 1. The results of our experiments for six different values of β, namely $\beta =$ 0.55, 1.1, 1.65, 2.25, 2.6, and 3.85 are tabulated below for three different values of m in Table I and graphed at two other values for m in Fig. 2 *a* ($m = 0.1$) and Fig. 2 *b* ($m = 0.3$). To determine X_A and X_B at any given value of m, say m_1, one ob-

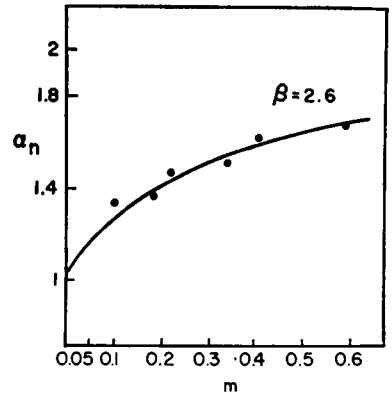


FIGURE 1 A smooth curve plotted through the raw data of α_n vs. m for $\beta = 2.6$.

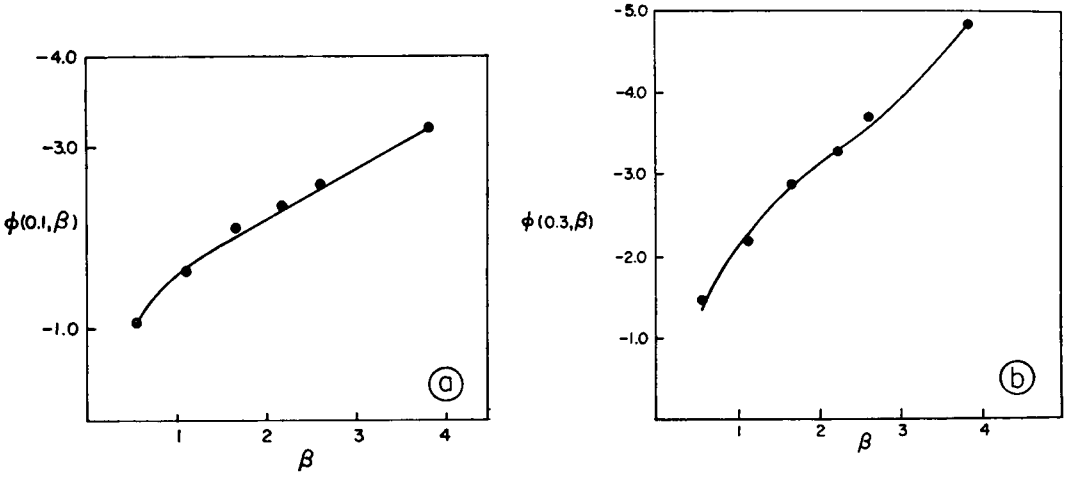
TABLE I
VALUE OF Φ (m,β) FOR THREE VALUES OF m

Φ (0.15, 0.55)	−1.19
Φ (0.15, 1.1)	−1.80
Φ (0.15, 1.65)	−2.22
Φ (0.15, 2.2)	−2.59
Φ (0.15, 2.6)	−2.92
Φ (0.15, 3.85)	−3.81
Φ (0.2, 0.55)	−1.27
Φ (0.2, 1.1)	−1.94
Φ (0.2, 1.65)	−2.40
Φ (0.2, 2.2)	−2.82
Φ (0.2, 2.6)	−3.21
Φ (0.2, 3.85)	−4.25
Φ (0.25, 0.55)	−1.35
Φ (0.25, 1.1)	−2.06
Φ (0.25, 1.65)	−2.56
Φ (0.25, 2.2)	−3.02
Φ (0.25, 2.6)	−3.45
Φ (0.25, 3.85)	−4.61

tains the derivative of $\Phi(m_1, \beta)$ at a certain value of β (Steiner, 1968). This yields $\ln X_B$ since

$$\left(\frac{d\Phi(m_1, \beta)}{d\beta}\right) = \ln X_B, \qquad (2)$$

and from $\ln X_B$ we obtain X_B and hence [B] or the monomer concentration. The derivatives in all cases were obtained by taking tangents to smoothed curves of $\Phi(m,\beta)$ vs. β. This procedure was checked in two cases against derivatives obtained



FIGURES 2 *a* and *b* In these figures the least-square polynomial fits are Φ (0.1, β) = $-0.0029564988 - 2.2219853\beta + 0.73070162\beta^2 - 0.096332943\beta^3$ and Φ (0.3, β) = $-0.0071562688 - 2.7920953\beta + 0.83903608\beta^2 - 0.11373051\beta^3$ in figs. 2 *a* and *b*, respectively. These graphs suggest that while polynomial fits can be used for interpolation they may not accurately reflect the functions beyond the range of the fit. Smooth curves through the data or spline function interpolations may be used for limited extrapolation.

TABLE II

VALUES OF [A] AND [B] AS A FUNCTION OF m

	β = 1.5		β = 2.0	
	[A]	[B]	[A]	[B]
	moles/liter	*moles/liter*	*moles/liter*	*moles/liter*
m = 0.1	0.044	0.048	0.034	0.055
m = 0.15	0.060	0.066	0.051	0.073
m = 0.2	0.074	0.082	0.066	0.088
m = 0.25	0.090	0.095	0.073	0.105
m = 0.3	0.105	0.108	0.081	0.123

from least-square polynomial fits at m = 0.1 and m = 0.3. The polynomials for those two cases are given in the legend of Fig. 2. There was little difference in the results of the two procedures. It must be emphasized that these polynomial fits are strictly for interpolation and should not be used for any extrapolative purpose. The value of $\ln X_A$ is obtained from equation 1, with $m = m_1$. Knowing $\ln X_A$ we then compute X_A and lastly the monomer concentration [A]. For two different values of β, namely β = 1.5 and β = 2.0, we have computed the values of [A] and [B] for m = 0.1, 0.15, 0.2, 0.25, and 0.3. These are shown in Table II. Knowledge of [A] and [B] at different values of m enables us to compute the equilibrium constants using another equation given by Steiner (1968), which is written as follows:

$$m - [A] - [B] - g([A]) - h([B])$$
$$= K_{11}[A][B] + K_{12}[A][B]^2 + K_{21}[A]^2[B] + \cdots, \quad (3)$$

where

$$g([A]) = \sum_{i>1} K_{i0}[A]^i,$$

and

$$h([B]) = \sum_{j>1} K_{0j}[B]^j. \quad (4)$$

It will be noted that equation 3 should hold for all values of β. Equation 3 can be simply interpreted as follows: We effectively subtract from m a series of terms corresponding to the monomer [A] and to the aggregates of A larger than the monomer, given by $g([A])$. Similarly we subtract a term for [B] and aggregates of B. This subtraction accounts for all species that are not in mixed aggregates and what is left are the species found in the mixed aggregates, which can be analyzed in terms of the series given by the right-hand side of equation 3.

The term $[A] + g([A])$ on the left-hand side of equation 3 is equivalent to the total molal concentration, m_A, of a solution of pure A, with no B present, when the concentration of monomeric A is equal to [A]. Similarly, the term $[B] + h([B])$ is equal to m_B, the total molal concentration of a solution of pure B when the concentration of monomeric B is equal to [B]. The left-hand side of equation 3 is thus equal to $m - m_A - m_B$, where m, m_A, and m_B are all obtained for the same values of [A] and [B].

The procedure adopted was as follows:

(*a*) Values of [A] and [B] were computed corresponding to a series of values of m and a particular value of β by the use of equations 1 and 2. For a single value of β one may thus tabulate values of [A] and [B] corresponding to each value of m.

(*b*) Separate measurements were performed upon solutions of pure A and of pure B. The mole fraction of monomeric A (and hence the molal concentration of free A monomers) when no B is present is given by (Steiner, 1954)

$$\ln X_A = \int_0^{m_A} (\alpha_n^{-1} - 1) \frac{dm_A}{m_A} \quad (5)$$

(when no B is present, $m_A = m$). In this way, values of [A] corresponding to a series of values of m_A were tabulated, as well as values of [B] corresponding to a series of values of m_B, so that m_A and m_B are known as functions of [A] and of [B], respectively.

(*c*) For a given value of m, m^*, the corresponding values of [A] and [B], [A*] and [B*], are known. Utilizing the data obtained for solutions of pure A and pure

TABLE III

VALUES OF [A] + g([A]) (OR m_A) AND [B] + h([B]) (OR m_B)

	β = 1.5		β = 2.0	
	[A] + g([A])	[B] + h([B])	[A] + g([A])	[B] + h([B])
m = 0.1	0.048	0.048	0.037	0.056
m = 0.15	0.065	0.067	0.055	0.077
m = 0.2	0.085	0.088	0.073	0.097
m = 0.25	0.103	0.105	0.083	0.120
m = 0.3	0.125	0.125	0.093	0.148

B, the values of m_A and m_B, m_A^* and m_B^*, which correspond to [A*] and [B*], are interpolated from the curves of m_A as a function of [A] and of m_B as a function of [B], respectively. The quantity $m^* - m_A^* - m_B^*$ is equal to the left-hand side of equation 3 when [A] = [A*] and [B] = [B*]. For inosine and cytidine the values of [A] + g([A]) (or m_A) and [B] + h([B]) (or m_B) corresponding to the monomer concentrations given in Table II are shown in Table III for β = 1.5 and β = 2.0. Returning to equation 4 after we subtract [A] + g([A]) and [B] + h([B]) from m, we are in a position to analyze the remainder of the expression in terms of the right-hand side of equation 3.

The data obtained for solutions of pure A and of pure B permit calculation of the consecutive association constants for the self-association of A and of B using the equations:

$$m_A = [A] + K_{20}[A]^2 + K_{30}[A]^3 + \cdots,$$

$$m_B = [B] + K_{02}[B]^2 + K_{03}[B]^3 + \cdots, \qquad (6)$$

where K_{20} and K_{02} are the equilibrium constants for dimerization of A and of B, etc. The numerical analysis was carried out by methods described elsewhere (Magar, 1969). For these conditions (H_2O, 50°C) the association constants for dimer formation (K_{11}) were found to be 1.1 $\pm$0.3 for inosine and 0.8 $\pm$0.2 for cytidine, with no assumptions as to the relative magnitude of the consecutive association constants. The details of the analysis for these and other nucleosides at various temperatures will be presented in a separate publication.[1] For comparative purposes the data for inosine were also analyzed in terms of the isodesmic model of Solie and Schellman (1968), which postulates that the stepwise association constants for each stage of association are equal. This approach yielded a value of 1.5 $\pm$0.5 for the association constant.

The fact that the right-hand side of equation 3 is an infinite series raised the question of how many terms one uses in the analysis. This question has been dealt

[1] M. E. Magar and R. F. Steiner. Manuscript in preparation.

with in some detail previously (Magar, 1969) and will be taken up again in the discussion. The variation of $m - m_A - m_B$ with [A] and [B] was analyzed in terms of two different models. If, as was done by Solie and Schellman (1968), a model of the isodesmic type is assumed, postulating a constant equilibrium constant, k, for each association process of the type

$$A + A_{i-1}\,B_j \rightleftarrows A_i\,B_j$$

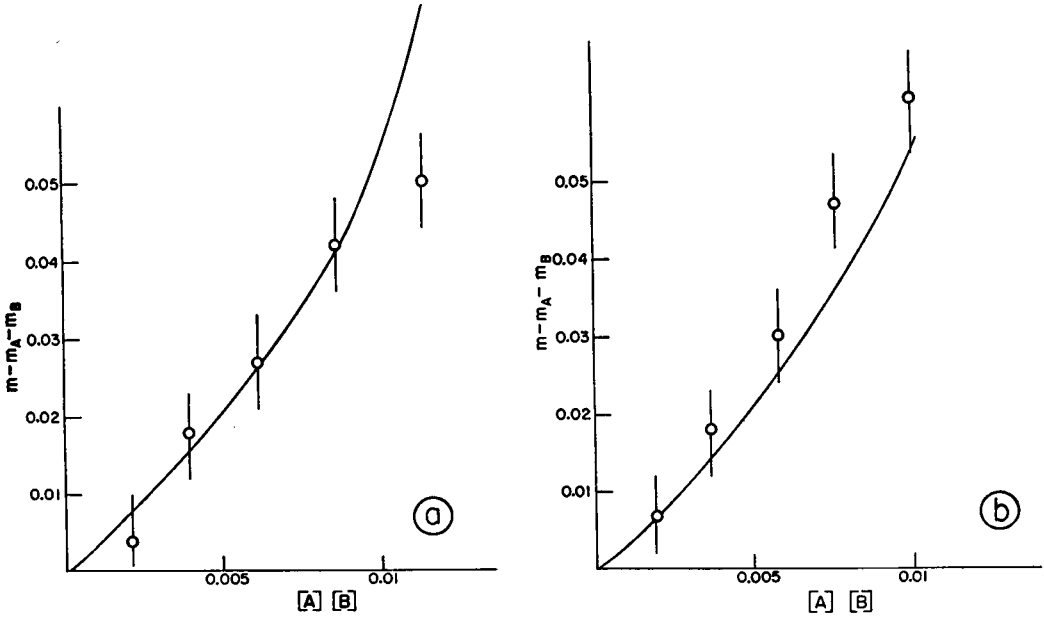
or

$$B + A_i\,B_{j-1} \rightleftarrows A_i\,B_j$$

then

$$\begin{aligned} m - m_A - m_B &= k\,[A][B] + k^2[A]^2[B] \\ &\quad + k^2[A][B]^2 + \cdots = \sum k^{i+j-1}[A]^i[B]^j \\ &= k[A][B]/(1 - k[A])(1 - k[B]). \end{aligned} \tag{6}$$

Within experimental uncertainty, the variation of $m - m_A - m_B$ with [A][B] could be fitted for $\beta = 1.5$ and for $\beta = 2.0$ by a value of k equal to 2.7 ± 1.3 (molal)$^{-1}$ (Fig. 3).



FIGURES 3 *a* and *b* A plot of $m - \{[A] + g([A]) + [B] + h([B])\}$ vs. [A][B] for $\beta = 1.5$ and $\beta = 2.0$. The curves are theoretical and are drawn using equation 7, assuming a value of k equal to 2.7.

Over this concentration range the data could also be formally fitted within experimental uncertainty on the assumption that only the species AB was present to a significant extent. This led to a value of $K_{11,app} = 4.9 \pm 2.0$ (molal)$^{-1}$. This model predicts a linear variation of $m - m_A - m_B$ with [A][B].

Because of the limited concentration range and the presence of experimental error, it is not possible to choose between the two models on the basis of these data. Since there is no obvious physical reason why only the first stage of association should be significant, the isodesmic model is probably closer to reality, although, if it is not strictly valid, the cited value of k will correspond to some kind of average. The alternative value of $K_{11,app}$ should probably be regarded as an upper limit, because of the nature of the assumed model.

Some discussion of the source and expected magnitude of errors is pertinent here. The quantities m and Φ can be obtained to within $\pm 2\%$. A chief source of error is in evaluation of $\ln X_B$ from equation 2. In the present case, where the variation of Φ with β shows little curvature in the region of interest, the estimated error in X_B or [B] is $\pm 7\%$. Because of the use of equation 1, errors in [A] and [B] will be of opposite signs; i.e. if [B] is too large, [A] will be too small and vice versa. The errors in m_A and m_B, which are not greatly different from [A] and [B] in the concentration range of interest, are likewise of the order of 7% and of opposite sign. Because of the latter factor the errors tend to cancel for the difference

$$m - m_A - m_B .$$

However, since the latter quantity represents the difference between two quantities of comparable magnitude, any residual error is magnified. The estimated final uncertainty is ± 1.3 for k and ± 2 for $K_{11,app}$.

The major result of this investigation is that cytidine and inosine undergo joint association. The fact that the same association constant was obtained for two different values of β enhances our confidence in the data.

The association of inosine and cytidine is a relatively unfavorable case for the application of theories of the association of mixed systems, since the self-association of the individual monomers is comparable in extent to that of the heterogeneous association of interest. This is primarily responsible for the relatively large magnitude of the experimental uncertainty.

The isodesmic model is of course only an approximation for the higher stages of association of a heterogeneous system such as the present one, since the formation of the complex species beyond AB may involve A-A and B-B linkages, as well as A-B linkages, and the free energies of association are unlikely to be identical for all three. The value of k cited should thus probably be regarded as an average.

DISCUSSION

To our knowledge, our value for the association constant for the joint interaction of inosine and cytidine is the first rigorous application of the theories of two solute

associations. Our primary purpose here is to illustrate the application of the theory to real data and to point out some of the problems encountered. However, since earlier work exists for similar systems, it is desirable to compare our approach with those of previous authors.

Solie and Schellman (1968) have made several rough estimates of association constants at 25°C of a number of deoxynucleoside mixed pairs. The estimates were made using equations pertaining to the theoretical treatment of single solutes only, as was recognized and stated by these authors. The approximate association constants given by Solie and Schellman (which we reproduce for the sake of comparison) for the following mixtures— dA-dT, dA-dC, dG-dC, dC-dT, and dA-dU[2] —are 3–6, 3–6, 4–8, 0.9, and 3, respectively. In each case the estimated range of values for the association constant was less than that for the corresponding purine-purine pair and intermediate to the values for the purine-purine and pyrimidine-pyrimidine pairs. However, the method of calculation employed can be expected to yield only a rough index of the order of magnitude. Moreover, the values must represent some kind of ill-defined average for the three kinds of nucleoside association occurring, since these authors had no means of differentiating quantitatively between the effects of self-association of the purine and pyrimidine nucleosides and those of the mixed association.

An alternative approach was developed in an early pioneering study by Ts'o et al. (1963, 1964) and more recently by Gratzer (1969). Both studies depended upon the solubilization of a sparingly soluble purine or pyrimidine base by a series of more soluble purine or pyrimidine nucleosides or nucleotides. Both found that purine-purine interactions were generally stronger than purine-pyrimidine, which in turn were stronger than pyrimidine-pyrimidine. An apparent exception was the case of the base purine itself, whose self-association constant, 2, (computed on the isodesmic model) was less than that of several mixed purine-pyrimidine pairs and approximately equal to that for the purine-thymine pair (Ts'o et al., 1963).

For convenience of analysis, these studies made use of the assumption that the association constant for the combination of the monomeric form of the base which was solubilized with a monomeric unit of the solubilizing nucleoside was not affected by the incorporation of the latter into a complex arising from self-association. In other words, if A represents the solubilized base and B, the solubilizing nucleoside, then each monomeric unit of the aggregate B_j combines with free monomeric A as if it were not self-associated.

No earlier data are available for the association of inosine with other nucleosides. However, the association constant for the self-association of inosine at 25°C is unusually small, being close to that for purine (Solie and Schellman, 1968).

The principal advantage of the method used in the present study is that it permits separation of mixed association from self-association. As was mentioned earlier,

[2] A, adenine; T, thymine; C, cytidine; G, guanine; U, uridine.

the data can, within experimental error, be fitted by either a single step or an isodesmic model, yielding association constants of 4.9 ±2 or 2.7 ±1.3, respectively. The constant for the self-association of inosine under the same conditions is 1.5 ± 0.5 for the isodesmic model and 1.1 ± 0.2 for the first stage of association, if the isodesmic assumption is not invoked.

It is the feeling of the authors that the isodesmic model is more plausible than the single step model for the mixed association and we shall present the remainlng discussion in terms of the association constant obtained using this model. The difference between this value and the association constant computed for inosine on the isodesmic model is within experimental error.

The magnitude of the experimental uncertainty involved in determining k for this system is too large to permit a decision as to the relative magnitude of the association constant for inosine and cytosine, as compared with the association constant for the self-association of the purine monomer. All that can be said is that they are of a similar order of magnitude.

The formation of hydrogen bonds between the bases hypoxanthine and cytosine cannot be excluded on the basis of the available data, but appears somewhat unlikely in view of the accessibility of the base pair to aqueous solvent. A more likely source of the stabilization energy of this mixed base pair is the stacking interaction of the planar bases.

While much information is available on the formation of the polyriboinosinic-polyribocytidylic helical complex (Chamberlain and Patterson, 1965; Davies and Rich, 1958; Miles, 1961; Felsenfeld and Miles, 1967; Sigler et al., 1962; Sarkar and Yang, 1965), any computation of the free energy of base stacking from such data involves too many uncertainties to make a comparison with our result useful.

The opinions in this paper are those of the authors and do not necessarily reflect the views of the Navy Department of the naval service at large.

The authors wish to thank Mr. Thomas J. Conner for modification of the instrument.

This work was Research Task MR005.06-0005A from the Bureau of Medicine and Surgery, Navy Department.

Received for publication 17 December 1969 and in revised form 16 February 1970 and 13 April 1970.

REFERENCES

ADAMS, E. T., A. H. PEKAR, D. A. SOUCEK, L. H. TANG, G. BARLOW, and J. L. ARMSTRONG. 1969 *Biopolymers.* **7**:5.

CHAMBERLAIN, M. J., and D. PATTERSON. 1965. *J. Mol. Biol.* **12**:410.

DAVIES, D. R., and A. RICH. 1958. *J. Amer. Chem. Soc.* **80**:1003.

FELSENFELD, G., and H. T. MILES. 1967. *Annu. Rev. Biochem.* **36**:407.

GRATZER, W. B. 1969. *Eur. J. Biochem.* **10**:184.

MAGAR, M. E. 1969. *J. Theor. Biol.* **22**:291.

MILES, H. T. 1961. *Proc. Nat. Acad. Sci. U.S.A.* **47**:79.

SARKAR, P. K., and J. T. YANG. 1965. *Biochemistry.* **4**:1238.

SIGLER, P. B., D. R. DAVIES, and H. T. MILES. 1962. *J. Mol. Biol.* **5**:709.
SOLIE, T. N., and J. A. SCHELLMAN. 1968. *J. Mol. Biol.* **33**:611.
STEINER, R. F. 1954. *Arch. Biochem. Biophys.* **49**:400.
STEINER, R. F. 1968. *Biochemistry.* **7**:2201.
TS'O, P. O. P., I. S. MELVIN, and A. C. OLSON. 1964. *J. Amer. Chem. Soc.* **87**:5241.
TS'O, P. O. P., I. S. MELVIN, and A. C. OLSON. 1963. *J. Amer. Chem. Soc.* **85**:1289.