Physical Properties of Some Ribosomal Proteins in Solution and Evidence for Molecular Interactions between Isolated Ribosomal Proteins†

Michael F. Rohde, Sally O'Brien, Susan Cooper, and Kirk C. Aune*

ABSTRACT: Many previous studies have been directed toward obtaining a physical visualization of the relationship between the protein and RNA in the ribosomal subunits isolated from *Escherichia coli*. The current study is the first report where an attempt has been made to directly assess interactions between a pair of isolated ribosomal proteins separate from the intact system by means of sedimentation equilibrium analysis. The molecular weights of the proteins S3, S4, S5, S6, S7, S8, and S20 from the 30S subunit of the *E. coli* ribosome were determined under *conditions of assembly of the subunit* by sedimentation equilibrium. All of the proteins exhibited molecular weights consistent with monomeric behavior (i.e., in agreement with the measurement of the ultimate molecular weight in denaturing solvents as reported in other studies as well as in the current study) except S8 which indicates a tendency to self-associate. Hydrodynamic measurements on the proteins indicate that these proteins are not completely disorganized in solution such as a random coil, although not as compact as globular proteins. The frictional coefficient ratios found for these ribosomal proteins range from 1.4 to 1.9. The hydrodynamic data are discussed as containing some evidence that stable interaction sites could exist in the proteins. The molecular weight data are considered pertinent to a sedimentation equilibrium study of protein–protein interactions that may be occurring in the ribosomal subunits. Two proteins, S3 and S5, considered in this investigation were found to exhibit no tendency to self-associate under conditions of reassembly. When the two proteins are mixed under those same conditions, however, a species with a molecular weight greater than that of either S3 or S5 is observed to be formed. The interpretation is presented that a molecular interaction between S3 and S5 is the cause. The system is described as containing S3, S5, and a complex between S3 and S5 with a stoichiometry of 1:1 and an association equilibrium constant of 5.7×10^5 l./mol ($\Delta G^\circ = -7.25$ kcal/mol). Since the association appears to be specific and of moderate strength, it is concluded that the interaction could have some pertinence with respect to conferring a structural arrangement in the ribosomal subunit. Moreover, it is concluded that protein–protein interactions, in general, must be considered in addition to the well documented significant RNA–protein relationships when models for ribosome structure and assembly are formulated.

The revelation that the 30S ribosomal subunit from *Escherichia coli* was the result of a systematic clustering of specific proteins and RNA was first reported by Mizushima and Nomura (1970). Since then many innovative studies have been reported attempting to describe in more detail the complex structure of the ribosome. The assembly scheme was further refined by other workers (Schaup et al., 1970; Craven and Gupta, 1970; Huang and Cantor, 1972; Bickle et al., 1972). The number (Schaup et al., 1971; Garrett et al., 1971) and composition (Zimmermann et al., 1972; Schaup and Kurland, 1972) of binding sites between 16S RNA and isolated ribosomal proteins have also been investigated. The picture that developed was one where a given protein was observed to bind to the RNA only when specific requirements were met. That is, in the case of the proteins S4, S7, S8, S15, and S20 the only requirement for binding was that RNA was present. In the case of the other 16 proteins shown to be present in the complete structure, there is a requirement for the prior presence of a specific protein from the above list before the additional proteins can be bound. The remaining proteins also exhibited a definite order in which binding could occur. This intricate network for assembly can be construed to mean more than the ability to observe binding in a definite order. It is clear that the specific order of binding can be the result of two effects: (1) that protein A causes the incomplete RNA–protein structure to take on a new conformation such that protein B is accommodated, and/or (2) that protein A directly interacts with B. Both types of binding effects occur in biological systems. The former might be considered to be related to the allosteric phenomenon observed in some polymeric enzyme systems. The latter simply involves direct interaction between two proteins where the interaction site is already in existence. Many examples of this type of interaction are also present in biological systems. The implication of an ordered scheme for assembly is that the latter type interaction may exist.

While the existence of protein–RNA interactions has been demonstrated, the importance of protein–protein interactions has been generally minimized (Kurland et al., 1972; Lutter et al., 1974); however, no report has been made of any attempt to detect or measure such interactions separate from the intact system.

In order for direct interaction between any two proteins to occur, an interaction site of a specific nature is assumed to be present. Specific sites on proteins are typically formed via the three-dimensional folding of proteins where the re-

† From the Department of Biochemistry, Ohio State University, Columbus, Ohio 43210, and the Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77025. *Received August 16, 1974.* This work was supported in part by the National Science Foundation, Grants GB-33623 and GB-40027.

* Address correspondence to Baylor College of Medicine.

quired side chains necessary for forming the site are brought together as a result of noncovalent interactions among the side chains comprising the protein. To be a site of any consequence biologically, it would not be unreasonable to assume that the site is a result of protein conformation which represents the predominant state out of all the possible states available to the protein. Such states are observed for proteins and are usually seen to be compact in nature as opposed to the extended structures that result when proteins are "denatured" by heat, extremes of pH, and exposure to solvents such as urea and Gdn · HCl[1] in water. Thus far almost all physical characterizations of the ribosomal proteins have been carried out in solvents such as 0.1% SDS, 6 *M* Gdn · HCl, and 6–8 *M* urea. The predominant state achieved in the absence of the preceding solutes would, most likely, be more closely related to that state found in the intact ribosome. If that state is involved in a specific interaction with another protein, it is possible that the interaction is observable isolated from the rest of the subunit. In order to study the *isolated pair interaction,* the properties of the *isolated* proteins must be available.

The argument proceeds that if one can demonstrate the isolated ribosomal proteins to have somewhat compact structures in solution, the possibility that specific sites of interaction are already present is demonstrated. This does not preclude the possibility of sites being created by an interaction for although a random coil does not possess a specific site (comprised of several side chains) by definition, it is conceivable that two random coils could sacrifice their random motion energy for a binding energy gained by interacting.

This study first considers whether the proteins investigated have hydrodynamic shapes, consistent with the concept that they possess structural integrity. If there is implied structural integrity it is not unreasonable to suggest that these structures also participate in the type of interaction which occurs in the formation of a fully assembled ribosome. There is some evidence for a few neighbor relations which have been implied by detection of chemically cross-linked pairs of ribosomal proteins (Bickle et al., 1972; Chang and Flaks, 1972; Lutter et al., 1974). This study also reports the detection and partial thermodynamic characterization of an interacting pair which has escaped detection by the chemical modification approach. Furthermore, the interaction is sufficiently strong to serve as an important driving force for the formation of the ribosome when several such interactions act in concert.

Materials and Methods

All pH measurements were made at room temperature using Radiometer pH meters, Models 4 and 26. Preparative centrifugation was done using either of the following Beckman centrifuges: L2, L3-50, or L2-65, with Beckman Type 30, 60 Ti, and Ti-14 rotors. Chromatographic columns were purchased from Pharmacia Fine Chemicals and Glenco Scientific. Column elutions were followed with an ISCO UA4 ultraviolet absorbance monitor. Conductivity measurements were made on a Yellow Springs Instruments conductivity meter, Model 31. Spectral readings were taken with a Cary 15 or Varian 635 spectrophotometer. Analytical ultracentrifugation experiments were conducted using four different Spinco Model E analytical ultracentrifuges, one of which was equipped with the Beckman photoelectric scanner and multiplexer. Optics were aligned for schlieren and Rayleigh interference work by standard procedures (Dyson, 1970; Richards et al., 1971) with camera lens focused at the two-thirds plane of the cell. For most of the experiments an AnD rotor was used, however, an AnG rotor was occasionally used with the scanner. Photographic plates were read in an LP-6 profile projector fitted with Nikon stage and micrometers. Amino acid analyses were performed on a Beckman amino acid analyzer, Model 116, according to the method of Spackman et al. (1958).

All chemicals used in this study were standard reagent grade. Phosphocellulose, purchased from Schwarz/Mann, had a 0.82 Mequiv/g capacity reported on the label. Gdn · HCl was purchased from Heico. SDS was Schwarz/Mann catalog 902696 (lauryl sodium sulfate).

Protein Purification. Frozen *E. coli* cells were purchased from General Biochemicals, Chagrin Falls, Ohio. Late-log, high peptone preparations of *E. coli* B (ATCC 11303) were used throughout this work. Ribosomes were obtained from the frozen cells by the procedure employed by Hindennach et al. (1971), with the exception that the thawed cells were disrupted using a French press at 10,000–15,000 psi. A yield of 2.5–3.5 g of ribosomes was obtained from each 250 g of the frozen cell preparation.

Separation into 30S and 50S subunits was accomplished by the procedure of Eickenberry et al. (1970), modified for use with a Beckman Ti-14 zonal rotor. The gradient was generated by the alternate method described by Eickenberry: buffered 45% sucrose was added to a 300-ml constant volume mixing chamber, initially containing buffered 7.4% sucrose, until 500 ml of gradient had been delivered to the rotor followed by a cushion of 45% sucrose. The optimal sample size for high yields of 30S ribosomes was 20 ml containing 600 mg of dissociated ribosomes. Overlay was 250 ml of buffer. Centrifugation was carried out for 3.5 hr at 48,000 rpm, at 5°. This procedure gave 100–150 mg of purified 30S and 250–300 mg of 50S ribosomes. Following dialysis to remove the sucrose, ribosomes were recovered by precipitation with two volumes of cold ethanol (Staehelin et al., 1969). The sucrose used for generating the gradients was treated with carboxymethylcellulose in the sodium form to remove possible RNase contamination (50 g of resin/l. of 50% unbuffered sucrose). Total 30S protein was extracted from the RNA with 67% acetic acid, then purified on cellulose phosphate columns in 6 *M* urea (Hardy et al., 1969). Buffered solutions of 6 *M* urea used for chromatography (Hardy et al., 1969) were usually prepared fresh. Any solution that had been stored in the cold longer than 2 weeks was analyzed on the amino acid analyzer for ammonia and methylamine. This procedure indicated that the change in methylamine concentrations was less than 1 m*M*, and in ammonia concentrations less than 6 m*M*; thus cyanate formation was not more than 6 m*M*. The reducing agent, 2-mercaptoethanol, was added to the buffers just prior to use and any solution containing 2-mercaptoethanol was discarded after 2 weeks. In order to obtain good separations of the proteins, prereduction with 2-mercaptoethanol was required. A 6 *M* urea solution of the mixed 30S proteins was adjusted to pH 8.4 and 2-mercaptoethanol added to make a 1% solution. After the mixture was stirred for 30 min at 25° the pH was adjusted to 5.6, then dialyzed at 4° vs. starting buffer for the column.

[1] Abbreviations used are: Gdn · HCl, guanidine hydrochloride; TMK, standard buffer containing 0.03 *M* Tris–0.02 *M* $MgCl_2$–0.35 *M* KCl (pH 7.4); SDS, sodium dodecyl sulfate.

Chromatography on phosphocellulose was carried out as described by Hardy et al. (1969).

Because of the large number of partially overlapping peaks that elute from the phosphocellulose columns and because the content of proteins in a "pure" 30S subunit need not be constant from preparation to preparation, the identification of the proteins in a given fraction can be difficult. Inital assignment as to protein content of the eluted fractions was based on a combination of information provided by the elution profiles from the phosphocellulose column and urea gel electrophoresis. Use of the ISCO column monitor at 280 nm gave very good correlation to the 230-nm elution pattern (Hardy et al., 1969), though some of the peak heights were different. Since the absorptivity of the proteins at 280 nm varies considerably, the ultimate decision of which tubes to pool was based on the bands observed in urea gel electrophoresis experiments on the column fractions. The urea gel was prepared containing 7.5% acrylamide, 0.8% methylenebisacrylamide, 8 *M* urea, and 0.3 *M* acetate at pH 4.3 (Bickle et al., 1973). The use of 8 *M* urea is preferred over 6 *M* urea for resolution of the bands, particularly in the case of the faster moving bands.

The pooled fractions were stored in either 15% acetic acid or 15% acetic acid containing 0.1 *M* KCl and 0.02 *M* $MgCl_2$. In either solvent, proteins could be maintained in solution up to 1 mg/ml. Higher concentrations often resulted in precipitation. The pooled fractions were analyzed by a modified SDS gel electrophoresis (Weber and Osborn, 1969) composed of 12.5% acrylamide and 1.25% bisacrylamide in a buffer containing 0.1% SDS and 0.1 *M* Tris phosphate (pH 7.2). Both SDS and urea gels were formed and run in a slab apparatus described by Reid and Bieleski (1968). After comparing the SDS and urea gels of the pooled fractions, preliminary assignments were made in correspondence to order of elution from the phosphocellulose column. Where necessary, rechromatography was performed in either 15% acetic acid, 0.1 *M* KCl on Sephadex G-100 columns, or on cellulose phosphate columns at a pH different from the original separation. A typical second separation consisted of applying 2–20 mg of protein to a 0.9 × 25 cm column, developed at 20 ml/hr using a 200-ml gradient covering a sodium chloride concentration range of 0.2 *M*. The higher flow rate permitted better recovery of protein than that recommended by Hardy et al. (1969). Proteins were considered to be purified when single bands were observed on both urea and SDS gels at a loading which would reveal 5% impurity. Amino acid analyses were in good agreement with data reported elsewhere (Craven et al., 1969; Kaltschmidt et al., 1970). The purified proteins were stored in the dry state or frozen in 15% acetic acid, 0.1 *M* KCl, and 0.02 *M* $MgCl_2$.

Once a protein had been purified in sufficient quantity its absorptivity was determined using the analytical ultracentrifuge (Aune and Timasheff, 1971). The absorptivities at 280 nm for total ribosomal proteins, S3 and S5 in 15% acetic acid were found to be 0.80, 0.88_8, and 0.34_3 ml/(mg cm), respectively. The calibration protein was lysozyme which has a dn/dC value of 1.70×10^{-4} l./g in that solvent. The proteins S4, S6, and S20 were found to have absorptivity values of 1.2_0, 0.9_9, and 0.28_2 ml/(mg cm) in TMK at 280 nm. Lysozyme exhibited a dn/dC value of 1.85×10^{-4} l./g in that solvent. The absorptivities were taken to be the same in both solvents for purposes of concentration determinations. The partial specific volumes were calculated from amino acid compositions (Cohn and Edsall, 1943) and densities were determined pycnometrically.

Molecular weights were determined by sedimentation equilibrium, employing the high speed method of Yphantis (1964). The majority of the molecular weight determinations were performed using Rayleigh interference optics. Experiments were routinely done using double sector cells with interference window holders, sapphire windows, and 12-mm Kel-F coated aluminum double sector centerpieces. The use of Teflon strips in place of the usual window cushion (Teller, 1973) generally reduced water blank corrections to less than 5 μ and correction was applied to the data only when deflection was significant.

Determinations obtained in later experiments using the photoelectric scanner at 280 nm to collect the data were in agreement with the interference method. Although the ultraviolet procedure yielded molecular weights that were essentially the same as those obtained from interference optics, there was greater scatter in the raw data which reflects the distribution of protein in the cell.

Solutions of protein in 6 *M* Gdn · HCl were prepared by dissolving lyophilized protein in 6 *M* Gdn · HCl at 0.1–0.4 mg/ml. The solution was dialyzed vs. at least two changes of the same solvent, allowing 12 hr or more for final dialysis equilibrium. The 6 *M* Gdn · HCl solutions were generally unbuffered and without reducing agent, unless amino acid analysis had indicated the presence of cysteine. For these proteins the solvent was modified to contain 0.1% 2-mercaptoethanol.

To obtain protein in the "native state," 0.1–0.4 mg of lyophilized protein was dissolved in 10–50 μl of 6 *M* Gdn · HCl. To this solution 0.5–1.0 ml of TMK buffer was added and the resulting solution incubated for 30–45 min at 37°. Dialysis was performed at 4° vs. TMK buffer with 3–5 changes, again allowing 12 hr or more for final dialysis equilibrium. Using this procedure precipitation of protein was minimal, though at concentrations of protein greater than 1.5 mg/ml, protein solubility became a definite problem.

Sedimentation velocity experiments were performed using double sector cells and the photoelectric scanner.

Calculation of Results. Photographic plates from sedimentation equilibrium experiments were aligned on the microcomparator stage, then the vertical position of three light fringes was measured and averaged. When the average and the central fringe were more than 5 μ apart, all three readings were repeated (Aune and Timasheff, 1971). This procedure was applied at 50–100-μ intervals, from a vertical displacement of 50 μ (relative to the fringes at the meniscus) to the base of the cell. For all experiments this gave an average of 30–40 data points.

The molecular weights of the single proteins were ascertained from sedimentation-equilibrium data by whole cell averages computed with a program developed in this laboratory (Kar and Aune, 1974), by a least-squares calculation of the slope of an $\ln f$ vs. r^2 plot, and by a nonlinear fit calculated by a modification of the direct search program. The latter fit minimizes the deviation between the experimental curve and a theoretical curve described by a single molecular weight and meniscus concentration. For heterogeneous systems of the type discussed here, the nonlinear approach provided a more interpretable form of fit leading to an analysis of the energetics of interaction.

The basic equation describing the observed concentration distribution in a sedimentation equilibrium experiment is

given in

$$f_{ij} = f_{aj} \exp(\sigma_j \Delta r_i^2/2) \quad (1)$$

where f_{ij} is the absolute concentration at the radial position r_i due to the jth component; f_{aj} is the meniscus concentration of the sedimenting species j; Δr_i^2 is the square of the radial position r_i minus the square of the meniscus radial position r_a. The quantity σ_j is defined in

$$\sigma_j = \frac{M_j(1 - \bar{v}\rho)\omega^2}{RT} \quad (2)$$

Here, M_j is the molecular weight of species j; $\bar{v}$ is its partial specific volume; ρ is the density of the solvent; ω^2 is the square of the angular velocity in radians per second; R is the gas constant; and T the temperature in degrees Kelvin. Equation 1 may be rewritten to describe the fringe displacements as measured by subtracting f_{aj} from both sides to give y_{ij}, the measured displacement at r_i due to the jth component relative to the displacement at the meniscus.

$$y_{ij} = f_{aj}[\exp(\sigma_j \Delta r_i^2/2) - 1] \quad (3)$$

For a mixture of species with molecular weight M_j, the total displacement y_i is represented by summing the right-hand side of eq 3 over all species:

$$y_i = \sum_j^S f_{aj}[\exp(\sigma_j \Delta r_i^2/2) - 1] \quad (4)$$

where S is the total number of species present in the system. In this equation all parameters are known or can be measured directly with the exception of the separate f_{aj}.

A set of f_{aj} yields an estimate of y_i. The residual at that point, δ_i, is the difference between the observed value y_i and the value calculated from the set of f_{aj}

$$\delta_i = y_i - \sum_j^S f_{aj}[\exp(\sigma_j \Delta r_i^2/2) - 1] \quad (5)$$

For the set of f_{aj}'s to be valid, all points on the observed curve should be reasonably approximated. An indication of the fit between theory and experiment is given by the average residual $\bar{R}$

$$\bar{R} = \sum_i^N \frac{|\delta_i|}{N - S - 1} \quad (6)$$

where N is the number of data points. The numerical value of $\bar{R}$ should correspond to the expected random error in the data collected; if it is larger than this, the set of f_{aj}'s does not adequately fit the data and new values of f_{aj} should be sought. The average residual need not be defined as in eq 6 but could be written in other forms. For example, the square root of the sum of the squares of δ_i might be used leading to a slightly different fit. The form presented was chosen because it gives equal weight to all data points and is not as sensitive as a squared form to deviations due to plate reading errors.

Any of several matrix reduction methods failed to yield physically meaningful results due to the near singularity of the matrices derived from the data. Therefore, a direct search iterative method was chosen (Hooke and Jeeves, 1961) to obtain the f_{aj}'s. This method proceeds by the following algorithm: (1) make an initial estimate of the desired f_{aj} and calculate $\bar{R}$; (2) generate a better value of f_{aj} by adding a change parameter to one of the f_{aj}'s; (3) when step 2 fails to decrease $\bar{R}$, make changes in two f_{aj}'s at once; (4) when step 3 fails, reduce the size of the change parameter and return to step 2. This simple method has performed very well in the case of three species with initial estimates of zero in step 1, proceeding to final values of $\bar{R}$ which are on the order of 5 μ, the expected error due to plate reading. The method can be programmed into a desk calculator and has usually reached convergence in 2–12 hr, requiring 300–2000 calculations of $\bar{R}$. The Hewlett Packard 9810 used for these calculations provided these limits because the step consuming the most time is the summation in $\bar{R}$ where many exponentials must be calculated rather than stored. Thus, in a sedimentation equilibrium experiment where a concentration distribution is achieved a mixture of two proteins that associate to form a 1:1 complex can be resolved in terms of the meniscus concentration of each species. These concentrations can be related to an association equilibrium constant. For the reaction

$$A + B \overset{K}{\rightleftharpoons} C \quad (7)$$

the association constant is written in terms of the molar concentrations

$$K = [C]/[A][B] \quad (8)$$

This expression is easily rewritten in g/l. concentrations by dividing molar concentrations by the molecular weights:

$$K = \frac{M_A M_B}{M_C} \frac{c_C}{c_A c_B} \quad (9)$$

The g/l. concentrations are directly related to fringe displacements through an optical constant, k, giving the equilibrium constant as a function of the individual fringe displacements at any point in the cell. At sedimentation equilibrium the equilibrium constant is obeyed at all points in the cell. Thus at the meniscus, where information about the individual species is obtained from the fit, sufficient information is available to determine the thermodynamics of the association.

$$K = \frac{M_A M_B}{M_C k} \frac{f_C}{f_A f_B} \quad (10)$$

The fitting procedure provides values of f_{aj} which are then used to calculate the equilibrium constant. The nature of the fitting procedure gives an equilibrium constant for the model assumed that is determined by all the data collected. Other models can easily be tested by providing the fitting program with different molecular weights.

Results and Discussion

The proteins investigated here were chosen because sufficient quantities could be obtained to carry out the necessary work and because assembly map work has suggested that some of the proteins might be proximal. This study used the procedure of Hardy et al. (1969) to initially separate the ribosomal proteins. Rechromatography of the proteins yielded the electrophoretic bands in a urea gel shown in Figure 1. The bands observed as contaminants of the main bands are sufficiently light to be considered unimportant with respect to the conclusions reported here. The separative properties under the conditions used were both charge and molecular weight. Although homogeneity is indicated under those conditions, the proteins were also run on SDS gels where single bands confirmed the molecular weight homogeneity implied from the urea gel results. Tentative name assignments were given to the proteins based on the position of chromatographic elution and the electrophoresis experiments.

These tentative assignments were confirmed by comparison of amino acid analysis to the published compositions of

Table I: Amino Acid Composition (mol %) of 30S Protein.[a]

Amino Acid	30S Protein						
	S3	S4	S5	S6	S7	S8	S20
Asp	7.5	8.4	8.5	11.0	7.5	7.0	8.6
Thr	3.9	3.3	5.1	5.7	3.3	5.1	3.8
Ser	4.9	5.7	4.8	4.0	4.6	5.6	7.7
Glu	10.1	11.8	9.4	16.0	12.0	9.7	7.0
Pro	4.0	3.7	3.6	2.0	3.5	3.2	1.0
Cys[b]							
Gly	8.4	8.5	12.3	5.8	6.6	8.2	2.4
Ala	9.8	8.9	12.5	11.3	12.8	9.4	15.5
Val	9.5	6.2	10.6	8.2	9.3	8.2	1.9
Met		0.7	2.8	3.5	1.9	3.0	4.1
Ile	7.2	4.0	6.1	4.5	3.7	4.4	8.8
Leu	6.7	10.2	6.6	4.5	7.3	5.8	5.4
Tyr	2.1	3.7	1.8	2.9	1.7	1.8	1.7
Phe	2.9	2.1	3.3	3.2	2.6	1.5	3.0
His	2.5	2.1	2.8	3.9	5.2	1.3	3.4
Lys	9.5	9.8	6.8	5.2	6.5	8.0	14.4
Arg	8.6	11.1	6.8	8.1	11.4	5.5	11.2

[a] Average of two 24-hr 6 *N* HCl hydrolysates at 110°. [b] Not determined.

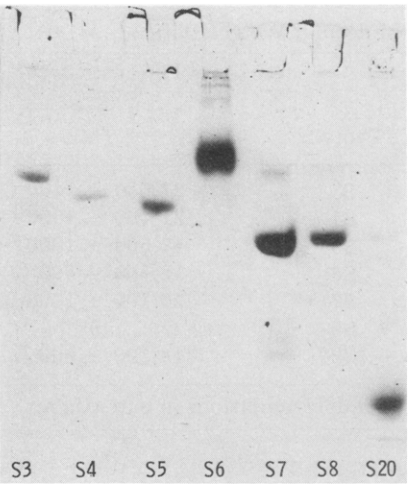

FIGURE 1: Urea gel electrophoresis of proteins. The figure illustrates the migration of proteins S3, S4, S5, S6, S7, S8, and S20 on urea gel prepared containing 7.5% acrylamide, 0.8% methylenebisacrylamide, 8 *M* urea, and 0.3 *M* acetate (pH 4.3). The gel was stained with Coomassie Blue.

Craven et al. (1969) and of Kaltschmidt et al. (1970). The compositions of the proteins studied here are presented in Table I. In order to assess the comparison of compositions in a somewhat quantitative manner, the following procedure was adopted.

The mole fraction of amino acid residue i in protein k, X_{ik}, was compared to the mole fraction of amino acid residue i in protein l, X_{il}. The discrepancies were quantitated by computing a correlation coefficient defined as

$$R_{kl} = \frac{\sum_i (X_{ik}X_{il})}{\left[\sum_i X_{ik}^2 \sum_i X_{il}^2\right]^{1/2}} \quad (11)$$

It is seen that a perfect correlation will yield a value of one. A totally imperfect correlation would yield a value of zero. In spite of the possible range of 0–1 for a correlation coefficient, proteins in general possess an intrinsic similarity in that most proteins contain some of each type of amino acid residue. The correlation coefficient computed by the above procedure therefore more typically ranges between 0.75 and 1.00. Table II contains a list of correlation coefficients for the proteins studied here vs. that published by Craven et al.

Table II: Correlation Coefficients.

Protein[a]	Protein						
	S3	S4	S5	S6	S7	S8	S20
S1	0.9650	0.9441	0.9595	0.9566	0.9400	0.9712	0.8089
S2	0.9625	0.9521	0.9672	0.9453	0.9400	0.9781	0.8741
S3	0.9938	0.9708	0.9646	0.9457	0.9648	0.9889	0.9012
S4	0.9730	0.9987	0.9395	0.9311	0.9641	0.9619	0.8701
S5	0.9661	0.9226	0.9967	0.9215	0.9421	0.9817	0.8283
S6	0.9070	0.9095	0.8932	0.9877	0.9497	0.9079	0.8345
S7	0.9716	0.9759	0.9465	0.9457	0.9843	0.9684	0.8834
S8	0.9843	0.9611	0.9786	0.9443	0.9544	0.9935	0.8772
S9	0.9683	0.9802	0.9423	0.9247	0.9581	0.9531	0.8610
S10	0.9470	0.9434	0.9154	0.9379	0.9478	0.9272	0.8310
S11	0.9362	0.9041	0.9531	0.8741	0.8889	0.9518	0.7867
S12	0.9576	0.9447	0.9344	0.8603	0.9280	0.9466	0.8164
S13[b]	0.9723	0.8076	0.9518	0.8365	0.8256	0.8093	0.9065
S14	0.9659	0.9804	0.9261	0.9058	0.9601	0.9502	0.9301
S15	0.9297	0.9708	0.9040	0.9086	0.9427	0.9289	0.8528
S16	0.9819	0.9539	0.9770	0.9461	0.9737	0.9665	0.8906
S17	0.9629	0.9365	0.9684	0.9175	0.9219	0.9725	0.8296
S18	0.9353	0.9539	0.8856	0.8930	0.9273	0.9061	0.9042
S19	0.9690	0.9572	0.9490	0.8948	0.9289	0.9664	0.8813
S20	0.9102	0.8958	0.8800	0.8807	0.9078	0.9073	0.9776
S21	0.9602	0.9647	0.9185	0.9049	0.9553	0.9464	0.8849

[a] Composition from Craven et al. (1969). [b] Composition from Kaltschmidt et al. (1970).

Table III: Molecular Weight Data.[a]

Protein	Sedimentation Equilibrium		SDS
	TMK	Gdn·HCl	Electrophoresis
S3	21,600 ± 1400 (9)	22,600 ± 500 (5)	25,100 ± 900 (7)
S4	21,400 ± 1000 (1)	22,400 ± 1000 (1)	23,500 ± 3000 (3)
S5	17,500 ± 1600 (8)	18,000 ± 2100 (13)	19,100 ± 700 (5)
S6	16,100 ± 500 (2)	15,700 ± 1600 (1)	15,900 ± 1600 (3)
S7	20,100 ± 1000 (2)	18,700 ± 1000 (1)	21,000 ± 2100 (2)
S8	[b]	16,600 ± 1200 (2)	15,800 ± 1600 (4)
S20	14,200 ± 800 (2)		17,000 ± 1700 (3)

[a] Number of determinations in parentheses. [b] Undergoes self-association.

Table IV: Frictional Properties of Proteins.[a]

Protein	$s_{20,w}$	f/f_{min}[b]	σ	f/f_{min}[c]
S3	2.28 ± 0.15 (2)	1.1_7	0.203 (1)	1.4_1
S4	1.65 ± 0.10 (2)	1.6_7	0.208 (1)	1.6_1
S5	1.49 ± 0.23 (4)	1.5_6	0.165 (2)	1.5_9
S6	1.75 ± 0.10 (2)	1.3_3	0.312 (1)	1.3_8
S7	1.66 ± 0.10 (2)	1.5_7		
S20	1.15 ± 0.05 (5)	1.8_8	0.209 (2)	1.8_3

[a] Number of determinations in parentheses. [b] Computed from eq 12. [c] Computed from eq 14.

(1969). An equally demonstrative table could be obtained by comparing a listing of correlations with the data of Kaltschmidt et al. (1970). The average correlation coefficient for the protein of one study vs. the other was 0.995 ± 0.003. It is seen that in each case the best correlation is obtained with the protein to which assignment had been made. In order to put the correlation coefficient into proper perspective, a calculation of the correlation coefficient of S7 vs. S7 with an imposed random error was performed. It was found that a random normal error of 1% of each residue produced correlation coefficients between 0.995 and 1.0. Above 1% random error the correlation coefficient dropped markedly, reaching a value of 0.96 when the error was 2%. The average correlation coefficient of background (the other 20 proteins) at that level of error was 0.913 ± 0.032, with one protein (S21) actually yielding a better correlation (0.975).

In order to derive frictional properties from hydrodynamic measurements of the proteins in solution, it is necessary to know the molecular weights of the proteins. The literature contains considerable variability in the molecular weights of several of the proteins, as indicated elsewhere (Dzionara et al., 1970; Craven et al., 1969). This study reports molecular weights of several of the proteins *in standard buffer* as determined by sedimentation equilibrium since this method will be utilized to detect protein–protein interactions. This fact is of particular importance for thus far only ultimate molecular weights of the proteins in denaturing solvents have been reported. In all cases except that noted, a homogeneous molecular weight distribution was observed in the experiments. These data are listed in Table III. It is also noted that excellent agreement is obtained for S4 where the primary structure is known (Reinbolt and Schiltz, 1973).

If these data reflect the ultimate molecular weight of the individual proteins, good agreement should be obtained with data obtained in Gdn·HCl solutions since Gdn·HCl should eliminate all noncovalent interactions (Tanford, 1968). The molecular weights uncorrected for preferential solvation are also listed in Table III. Good agreement is noted where comparisons can be made.

An estimate of the ultimate molecular weight of proteins can also be obtained utilizing the method of SDS electrophoresis as described by Weber and Osborn (1969). Using markers of carbonic anhydrase, chymotrypsinogen, β-lactoglobulin, lysozyme, and α-chymotrypsin chains on 12.5% gels with 1.25% cross-linking, the data in Table III were obtained for the ribosomal proteins. It is noted that some discrepancies occur between the thermodynamic measurements and the electrophoretic method. The most notable discrepancy is protein S3 which by sedimentation equilibrium has a molecular weight of 21,600, but by SDS electrophoresis has a molecular weight of about 25,100. No adequate explanation is presently available for such a deviation. It should be noted, however, that an unusual amino acid sequence could possibly alter the radius of gyration sufficiently to give rise to a more slowly migrating electrophoretic species.

The molecular size and shape of the ribosomal proteins in standard buffer were assessed by sedimentation velocity and in a few cases by gel permeation chromatography. The solubility of most of the ribosomal proteins appears somewhat limited, for refolding procedures yielded protein concentrations in standard buffer rarely above 1.5 mg/ml. Subsequently, the sedimentation coefficients reported in Table IV are considered to be reasonably close to infinite dilution such that no interpretation errors are made as a result of the concentration dependence of $s_{20,w}$.

The frictional coefficient ratio, f/f_{min}, is the ratio of the observed frictional coefficient to the minimum frictional coefficient where it is assumed that the molecule is an unhydrated sphere. Combination of the equation which defines the minimum frictional coefficient, f_{min} (Tanford, 1961), and the sedimentation coefficient relationship leads to the equation

$$f/f_{min} = \frac{(4/3)^{1/3}}{6\eta(\pi N)^{2/3}}\left(\frac{1-\bar{v}\rho}{(\bar{v})^{1/3}}\right)\frac{M^{2/3}}{s} \qquad (12)$$

where η is the viscosity of the solvent; ρ, the density of the solvent; N, Avogadro's number; $\bar{v}$, the partial specific volume of the protein; M, the molecular weight; and s, the sedimentation coefficient. These calculated values are also

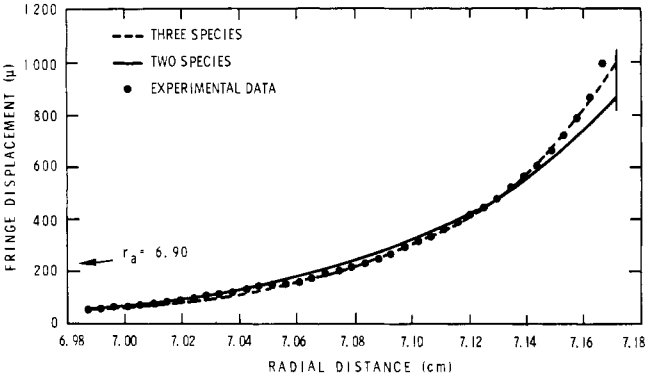

FIGURE 2: Sedimentation equilibrium pattern of S3–S5 mixture in TMK buffer. Rotor speed 28,000 rpm; temperature 5°; column height 2.7 mm; initial protein concentration 0.125 mg/ml of S3 and 0.125 mg/ml of S5. (●) Experimental points; (- - -) the theoretical fit for the three species S3, S5, and S3-S5 dimer; (—) the best obtainable fit for the two species S3 and S5.

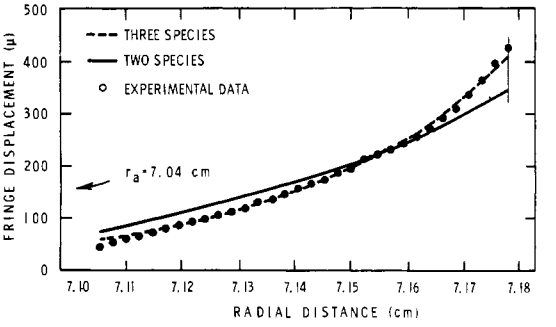

FIGURE 3: Same as Figure 2 with the following exceptions: rotor speed 32,000 rpm; column height 1.3 mm; initial protein concentration 0.060 mg/ml each of S3 and S5.

found in Table IV. To place these values in proper perspective, typical values of f/f_{min} for globular proteins are on the order of 1.20–1.30 while f/f_{min} for chymotrypsinogen, a random coil in Gdn · HCl (Tanford et al., 1967), is calculated to be 2.35.

A measure of f/f_{min} can also be obtained from gel permeation chromatography assuming that there are no interactions between protein and the gel matrix. The treatment of Ackers (1967) is used where the Stokes radius, R_s, is related to the partition coefficient, σ, through the function

$$R_s = B + A \text{ erfc}^{-1}(\sigma) \quad (13)$$

where A and B are constants.

A calibration of a G-75 column using known proteins leads to a calculation of R_s for the ribosomal proteins. Then f/f_{min} is calculated from the equation

$$f/f_{min} = R_s/R_{min} \quad (14)$$

where R_{min} is again calculated from the partial specific volume and molecular weight. These values are also listed in Table IV where good agreement is obtained between sedimentation and gel permeation analyses.

The frictional coefficient ratios reported in Table IV indicate that the ribosomal proteins, with the exception of S3, are somewhat more expanded than typical globular proteins, but not grossly unfolded as proteins are commonly found to be in 6 M Gdn · HCl. It is also seen in Table IV that the "core proteins," S4, S7, and S20, have frictional coefficient ratios greater than 1.5 while of the three "split proteins" studied only S5 has a frictional coefficient greater than 1.5. Although this observation is of limited significance presently, it is conceivable that this may be related to the amount of the conformational stability which the core proteins could receive through direct RNA interactions. Alternatively, the extended structure of the core proteins, as revealed from the hydrodynamic data, could be indicative of the structure possessed by the proteins when bound to the polyanionic RNA.

A direct test of protein–protein interaction can be achieved by conducting sedimentation equilibrium experiments on mixtures. The results of this study provide the molecular weights of the isolated proteins. A reliable measure of the molecular weight is required for mixture analysis. The values obtained by sedimentation equilibrium for proteins in TMK buffer were used for all further calculations and is assumed to be the most reliable, especially when considering the general agreement with the values obtained in 6 M Gdn · HCl. This assumption is made because molecular weights determined in the centrifuge are based on rigidly controlled physical parameters (of speed, temperature, and solution density) and are not influenced by the geometry of the molecule in solution.

In all cases, the sedimentation equilibrium experiments in 6 M Gdn · HCl provided linear plots of $\ln f$ vs. r^2 for single proteins. The experiments with the same proteins in TMK also provided linear plots; *this is taken as strict evidence that no self-association was detectable for these proteins in TMK buffer.* Two samples of S3 which had been stored in TMK buffer in the refrigerator for periods exceeding 2 weeks showed a curvature in the $\ln f$ vs. r^2 plot. Point molecular weight averages calculated on these two samples indicated the presence of a larger species than the monomeric S3. This condition was not easily reproducible and was not observed in freshly refolded samples of protein. All subsequent experiments were completed within 3 days of refolding the proteins in order to avoid this apparent aggregation.

When sedimentation equilibrium experiments are performed with mixtures of S3 and S5, plots of $\ln f$ vs. r^2 are curved as would be expected for a heterogeneous sample. The presence of species with molecular weights higher than either of the individual proteins was indicated upon observation of the calculated point molecular weight averages. If only unassociated S3 and S5 were present in the system, all molecular weight averages would fall between the molecular weight of the two species. In order to account for the increase in molecular weight, a fit of the data was calculated by means of the procedure outlined for three species present: S3, S5, and a 1:1 complex of S3 and S5. The results of these fits were very good; the three meniscus concentrations calculated by the direct search program described the experimental curves with an average residual, generally in the range of 2–6 μ, which is well within the error expected from plate reading scatter. As a further demonstration that the data could not be accounted for by assuming the lack of an interaction, fits were calculated allowing only the two monomeric species. The average residual for these calculations ranges between 20 and 30 μ and is too large to be explained as simply a random data collection error. This is emphasized in both Figures 2 and 3, where the solid lines describe the fit in terms of two species. The excellent agreement with experimental data in terms of a three species fit is illustrated by the dashed lines. The largest deviations in these fits are seen near the base. Such deviations can be ascribed

Table V: S3–S5 Complex Formation.

Association constant	$K = 5.7 \times 10^5$ l./mol $\pm 1.0 \times 10^5$
Free energy of association	$\Delta G^\circ = -7.25$ kcal/mol ± 0.56

to inaccuracies of plate reading since in taking readings where the fringes rise steeply and can often become less distinct, such an error is not unexpected. However, the systematic deviations from the two species curves in both figures are far beyond the reasonable limits of the method.

Yet another argument in favor of the complex comes from a consideration of conservation of mass. Knowing the meniscus concentrations of each species, it is a simple matter to calculate the total amount of each species present in the cell. The experiments were initialized with equal weight concentrations of the two proteins. The two-species fit in Figure 2 yields a mass ratio of S3 to S5 in the cell of 9.35; for the three species fit, the mass ratio of S3 to S5 from all species is computed to be 1.04. Similar results were obtained for all other experiments. These results are especially gratifying because the method does not apply this constraint, rather it is derived as a result only by requiring that the three f_{aj} fit the observed data.

An association constant was calculated from the three meniscus concentrations by eq 10 for each experiment. Table V presents an average value for nine experiments. These determinations were made using a variety of rotor speeds, column heights, and initial loading concentrations to produce the average reported.

Also presented in Table V is the average Gibb's free energy change calculated from each of the nine association constants for the S3–S5 interaction in TMK buffer at 5°.

In summary, this study investigated the hydrodynamic behavior of some of the proteins in the 30S ribosomal subunit from *E. coli*. It has been shown that although the proteins do not exhibit complete compactness in an isolated condition, they do not necessarily have the hydrodynamic properties of a completely disorganized polymer. In addition, the molecular weight analysis of the proteins under native and denatured conditions shows that all proteins studied excluding S8 exhibited no self-association under native conditions, a fact pertinent to the results for mixture of S3 and S5.

In view of the data presented here for that mixture it is clear that interactions between ribosomal proteins should not necessarily be eliminated when considering models of ribosome structure and assembly. The measured interaction between proteins S3 and S5 is considered to be significant for the following reasons.

(1) Interaction is always detected immediately after mixing the two proteins, and demonstrates no short term time dependence which could significantly alter the interpretation. Although nonspecific aggregation is seen in some samples of S3, no particular problem is encountered over the time course of the experiments carried out here. Thus, the state of the protein refolded under the described conditions contains sufficient structure that interaction between these two proteins does not require the binding energy of further components (i.e., RNA) in order to form a specific complex. This interaction was observed whether the proteins were refolded separately then mixed, or refolded together—in concert with the concept that the proteins can find their final state independent of other components provided sufficient thermal energy is available.

(2) Since pure isolated proteins studied here show no self-association under the same conditions, the interaction must be of a heterogeneous nature as would be found in the least complicated model—a S3–S5 dimer. "Goodness of fit" alone cannot rule out a higher order complex, but this would appear to be highly improbable in the light of stoichiometry measurements for these two proteins (Voynow and Kurland, 1971; Weber, 1972). These two proteins, and in fact all 30S proteins, have been reported to be present in amounts that correspond to one or less copies per ribosome.

(3) The value of the equilibrium constant does not vary with total protein concentration or reasonable length of time after refolding. This provides the information that all protein molecules are entering into the complex and that the proposed model, of one S3 and one S5 combining to form a 1:1 complex, is correct. Under the conditions used a higher order complex such as 1:2 or 2:2 is not consistent with the sedimentation equilibrium data, for it would give rise to a variation in the equilibrium constant of 1–2 orders of magnitude and about 1 order of magnitude in the mass ratio calculated on the assumed model. This indicates that no higher order complex can be detected by the method and that the predominant complex is the postulated 1:1 S3–S5 dimer.

(4) The free energy difference is large and negative on the order of the difference between the native and denatured states of many other proteins (Brandts, 1964; Aune and Tanford, 1969; Salahuddin and Tanford, 1970). This then represents a moderately strong and specific affinity between two proteins which have been previously implicated to influence each other in binding to a reconstituted ribosome (Mizushima and Nomura, 1970). The fact that this interaction alone is not strong enough to maintain structural integrity, as is seen in the whole ribosome, *indicates the possible interdependencies that could exist* between the moieties comprising the structure of the ribosome. Evidence is rather compelling that a *combination* of additional protein–protein *and* protein–RNA interactions does in fact lead to the complex that is known as the ribosome. Direct interactions between single 30S protein and 16S RNA have been confirmed for only five proteins; the interactions between the other 16 proteins and the RNA have been presumed to occur only when one or more of the five are present. Arguments have been presented that binding of the first proteins induces conformational changes which uncover new sites on the RNA (Lutter et al., 1974). Such a constraining model may be unnecessary if each component of the system participates in two or more interactions of the type seen between S3 and S5. A multiplicity of such interactions could produce a complex with the observed degree of structural unity found in whole ribosomes.

Acknowledgment

We thank Michael Lee for providing a Fortran copy of the direct search program and for helpful suggestions for its use.

References

Ackers, G. K. (1967), *J. Biol. Chem. 242,* 3237.

Aune, K. C., and Tanford, C. (1969), *Biochemistry 8,* 4579.

Aune, K. C., and Timasheff, S. N. (1971), *Biochemistry*

10, 1609.
Bickle, T. A., Hershey, J. W., and Traut, R. R. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 1327.
Bickle, T. A., Howard, G. A., and Traut, R. R. (1973), *J. Biol. Chem. 248,* 4862.
Brandts, J. F. (1964), *J. Am. Chem. Soc. 86,* 4291.
Chang, F., and Flaks, J. G. (1972), *J. Mol. Biol. 68,* 117.
Cohn, E. J., and Edsall, J. T. (1943), Proteins, Amino Acids and Peptides, New York, N.Y., Reinhold, p 375.
Craven, G. R., Voynow, P., Hardy, S. J. S., and Kurland, C. G. (1969), *Biochemistry 8,* 2906.
Craven, G. R., and Gupta, V. (1970), *Proc. Natl. Acad. Sci. U.S.A. 67,* 1329.
Dyson R. (1970), *Anal. Biochem. 33,* 193.
Dzionara, M., Kaltschmidt, E., and Wittmann, H. G. (1970), *Proc. Natl. Acad. Sci. U.S.A. 67,* 1909.
Eickenberry, E. F., Bickle, T. A., Traut, R. R., and Price, C. A. (1970), *Eur. J. Biochem. 12,* 113.
Garrett, R. A., Rak, K. H., Daya, L., and Stöffler, G. (1971), *Mol. Gen. Genet. 114,* 112.
Hardy, S. J. S., Kurland, C. G., Voynow, P., and Mora, G. (1969), *Biochemistry 8,* 2897.
Hindennach, I., Stöffler, G., and Wittmann, H. G. (1971), *Eur. J. Biochem. 23,* 7.
Hooke, R., and Jeeves, T. A. (1961), *J. Assoc. Comput. Mach. 8,* 2.
Huang, K., and Cantor, C. R. (1972), *J. Mol. Biol. 67,* 265.
Kaltschmidt, E., Dzionara, M., and Wittmann, H. G. (1970), *Mol. Gen. Genet. 109,* 292.
Kar, E. G., and Aune, K. C. (1974), *Anal. Biochem. 62,* 1.
Lutter, L. C., Bode, U., Kurland, C. G., and Stöffler, G. (1974), *Mol. Gen. Genet. 129,* 167.
Mizushima, S., and Nomura, M. (1970), *Nature (London) 226,* 1214.
Reid, M. S., and Bieleski, R. L. (1968), *Anal. Biochem. 22,* 374.
Reinbolt, J., and Schiltz, E. (1973), *FEBS Lett. 36,* 250.
Richards, E. G., Teller, D., and Schachman, H. K. (1971), *Anal. Biochem. 41,* 189.
Salahuddin, A., and Tanford, C. (1970), *Biochemistry 9,* 1342.
Schaup, H. W., Green, M., and Kurland, C. G. (1970), *Mol. Gen. Genet. 109,* 193.
Schaup, H. W., Green, M., and Kurland, C. G. (1971), *Mol. Gen. Genet. 112,* 1.
Schaup, H. W., and Kurland, C. G. (1972), *Mol. Gen. Genet. 114,* 350.
Spackman, D., Stein, W. H., and Moore, S. (1958), *Anal. Chem. 30,* 1190.
Staehelin, T., Maglott, D., and Munro, R. E. (1969), *Cold Spring Harbor Symp. Quant. Biol. 34,* 39.
Tanford, C. (1961), Physical Chemistry of Macromolecules, New York, N.Y., Wiley, p 357.
Tanford, C. (1968), *Adv. Protein Chem. 23,* 121.
Tanford, C., Kawahara, K., and Lapanje, S. (1967), *J. Am. Chem. Soc. 89,* 729.
Teller, D. (1973), *Methods Enzymol. 27,* 346.
Voynow, P., and Kurland, C. G. (1971), *Biochemistry 10,* 517.
Weber, H. J. (1972), *Mol. Gen. Genet. 119,* 233.
Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244,* 4406.
Yphantis, D. A. (1964), *Biochemistry 3,* 297.
Zimmermann, R. A., Muto, A., Fellner, P., Ehresmann, C., and Branlant, C. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 1282.