The Ribosomal Proteins of *Escherichia coli*. II. Chemical and Physical Characterization of the 30S Ribosomal Proteins*

G. R. Craven, P. Voynow,† S. J. S. Hardy, and C. G. Kurland

ABSTRACT: Amino acid and tryptic peptide analyses were performed on each of the 21 30S ribosomal proteins, whose purification is described in the previous paper; 18 of the proteins are chemically unrelated by these criteria. Two proteins are very similar. One other has not been sufficiently well defined to eliminate the possibility that it is a fragment of another purified protein.

In addition, there may be one protein in the 30S particle that we have not purified. We conclude that there are between 19 and 22 proteins in the 30S ribosomal subunit of *Escherichia coli*. The 30S ribosomal subunits used as the starting material in this work contain between 230,000 and 280,000 daltons of protein. The sum of the molecular weights of the 19 proteins, measured by high-speed equilibrium centrifugation, is 410,000 daltons. Therefore one copy of each protein cannot be accommodated by a single 30S particle. We tentatively conclude that the 30S particles of *E. coli* are heterogeneous.

The main objective of the present study is to determine the number of different proteins in the 30S subunit of *Escherichia coli*. This is a necessary step in establishing whether or not the ribosomes are homogeneous, and can be accomplished by demonstrating that each nominal ribosomal protein is a single, chemically distinct entity. Our first step in such an analysis is described in the preceding paper (Hardy *et al.*, 1969); 21 electrophoretically and chromatographically unique proteins have been purified from the 30S ribosomal subunit. We now describe the preliminary chemical and physical characterization of these proteins.

All 21 proteins that were isolated from the 30S subunit are chemically distinct. However, some ambiguities in the data have limited our confidence to the positive identification of only 19 of these as unique ribosomal proteins. There may be three more proteins in the 30S subunit. The molecular weight measurements of the identified proteins permit us to begin an analysis of the distribution of these proteins among the 30S particles. The combined molecular weight of the nineteen proteins is much greater than the average mass of protein per 30S particle. This suggests that the ribosomes may be heterogeneous.

Materials and Methods

Amino Acid Composition. The protein fractions from the phosphocellulose chromatography were concentrated using a Diaflo (Amicon Corp., Lexington, Mass.) ultrafiltration membrane, series UM-2. The concentrated solutions were then dialyzed against distilled water and aliquots taken for analysis. The aliquots (between 0.1 and 0.4 mg) were lyophilized in a heavy-wall Pyrex glass tube after which 0.05 ml of constant-boiling redistilled HCl (approximately 5.7 N) was added. The tube was constricted and thoroughly evacuated for 15 min. The evacuated tube was sealed and placed at 110° for 18 hr. The hydrolysate was taken rapidly to dryness under vacuum and analysis was performed by a modification of the method of Spackman *et al.* (1958). The analyzer used was a Spinco Model 120 C, supplied with UR-30 resin for accelerated analyses. Total analysis time with this system is 130 min.

Trypsin Digestion. The pooled protein fractions were dialyzed against distilled water and then adjusted to 0.2 M NH_4HCO_3 (pH 8.6). The protein concentration varied from 0.5 to 0.05 mg per ml, the total amount of protein usually being between 1.5 and 3.0 mg. The protein was digested for 4 hr at 37° with trypsin (Worthington Biochemical Corp., Freehold, N. J.) treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone. Approximately 0.25 mg of trypsin was used for each digestion. After digestion was complete, the sample was lyophilized.

Peptide Mapping. The technique used in these studies for peptide mapping has been recently reviewed by Hill and Delaney (1967) and is a modification of the approach used by Jones (1964). The trypsin digest is dissolved in a 0.2 M pyridinium acetate buffer pH 3.1 and adsorbed to Dowex 50X-8 resin (Spinco type PA-35) equilibrated with the same buffer. The column is maintained at 56° and is eluted with a linear gradient of pyridinium acetate going from 0.2 to 2.0 M and from pH 3.1 to 5.0. The effluent is directed into the ninhydrin reaction and analysis system of a Spinco 120 C amino acid analyzer. The gradient elution is maintained for 325 min at a flow rate of 70 ml/hr. At 325 min a timed motovalve is actuated which switches from the gradient chambers to a chamber containing only the limiting pyridinium acetate buffer. The limiting buffer elution proceeds for an additional 125 min at which time the analysis is terminated. The column dimensions are 0.9 × 23 cm. The ninhydrin flow rate is also 70 ml/hr.

Cyanogen Bromide Digestion. The conditions used for cyanogen bromide digestion were essentially those previously

* From the Department of Zoology, Department of Genetics, and Laboratory of Molecular Biology, University of Wisconsin, Madison, Wisconsin 53706. *Received February 13, 1969.* This research was supported by U. S. Public Health Service Grants GM 13832, GM 12411, and 5-PO1-GM 15422, as well as by funds from the Wisconsin Alumni Research Foundation.

† Predoctoral fellow of the National Institutes of Health.

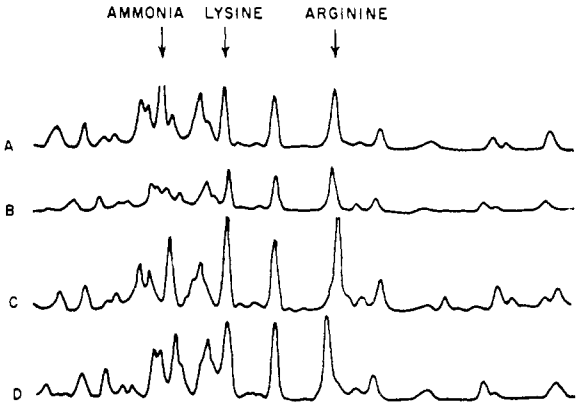

FIGURE 1: Dowex 50 separation of tryptic digests of four independent preparations of protein 9. Pattern B is from protein 9 extracted from K strain; all others are from B strain *E. coli*.

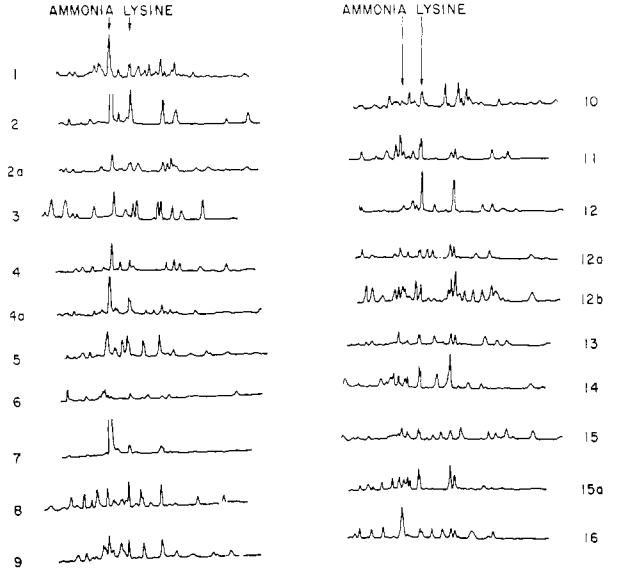

FIGURE 2: Comparison of tryptic peptide patterns of all purified 30S ribosomal proteins. The patterns are aligned by the chromatographic positions of ammonia and lysine.

described (Steers *et al.*, 1965). The protein was treated with a 50–100-fold excess of cyanogen bromide (Eastman Organic Chemicals) in a 70% formic acid solution. The digestion proceeded at room temperature for 20 hr and then was lyophilized. The lyophilized digest was dissolved in 8 M urea and used directly for polyacrylamide gel electrophoresis analysis.

Radioactivity. Water-containing samples were dissolved in scintillation fluid containing toluene, 5 g/l. of 2,5-diphenyloxazole, and 60 ml/l. of Biosolve BBS-2 (Beckman Instruments). A Beckman Model LS-250 ambient temperature scintillation spectrometer was used.

When radioactive proteins were analyzed the effluent stream was split into two equal portions by the use of an additional pump. One stream was analyzed by the analyzer's ninhydrin system and the second was fed to a Spinco Model 132 fraction collector. The fractions were lyophilized over NaOH pellets and H_2SO_4, dissolved in 1 ml of water and counted.

Equilibrium Sedimentation Analysis. All sedimentation runs were made in a Spinco Model E ultracentrifuge with RTIC and either an electronic or mechanical speed control. The An-D rotor containing cells equipped with sapphire windows and a six-channel centerpiece, which permits the simultaneous analysis of three solution–solvent pairs, was used in conjunction with the Rayleigh optical system. Each channel of the centerpiece was filled with 0.1 ml of liquid; this gives a column height of approximately 3 mm. It is necessary to apply a thin film of (Beckmann vacuum) grease to the faces of the centerpiece in order to prevent leakage of the cell at speeds in the neighborhood of 50,000 rpm. The interference patterns were photographed on type II G spectroscopic plates (Eastman Kodak) and were analyzed with a microcomparator (Gaertner Scientific Corp.). Fringe positions were measured by averaging the vertical position of five successive fringes at each horizontal position that was chosen.

Samples of between 0.5 and 1.5 ml were dialyzed for 5–15 days at 4° against 500 ml of standard urea buffer, pH 6.5 (Hardy *et al.*, 1969), to which was added 0.75 ml of β-mercaptoethanol just prior to dialysis. The dialysis flasks were sealed with Parafilm and stirred throughout dialysis.

Apparent molecular weights were calculated using values for ρ of 1.097 and 1.122 for standard 6 M urea buffer and 5 M guanidine hydrochloride, respectively (Kawahara and Tanford, 1966). The $\bar{v}$ of each protein was calculated from its amino acid composition (see Table III) by the method described by Schachman (1957). The values we have used are in good agreement with previously measured values of 0.740 and 0.734 (Möller and Chrambach, 1967) for unfractionated ribosomal protein.

The molecular weights of the proteins were measured by the high-speed, short-column method of Yphantis (1964). The samples were centrifuged for 24–50 hr at 4° and at approximately 48,000 rpm; protein 1 was run at 30,000 rpm because of its large size. Only fringe displacements greater than 100 μ ($\sim^1/_3$ fringe) were plotted. Most of the plates yielded at least ten usable points.

Some molecular weight measurements were also made by the low-speed equilibrium technique that employs the white-light fringe method of Richards *et al.* (1968). However, with solutions of 6 M urea or 5 M guanidine hydrochloride unusual precautions are required for the synthetic boundary runs. First, in order to avoid evaporation from or condensation into both the sample and solvent, they were transferred from dialysis to the centrifuge cell extremely rapidly and in the cold. If this is not done, a slight difference in urea concentration between sample and solvent can be established; this results in enormous errors in the synthetic boundary measurement. Second, once the loaded cell was placed in the chilled rotor, the rotor was shaken for about 1 min and then brought to the desired speed in the centrifuge without waiting for the evacuation of the centrifuge chamber. If this is not done, the fringes formed above and below the boundary are wildly curved. The shaking possibly dissolves urea crystals which may have formed on the cell walls. Such crystals can create local concentration inhomogeneities unless they are dispersed before the run begins.

Results

Amino Acid Compositions. The amino acid compositions of the 21 purified proteins expressed as mole per cent are sum-

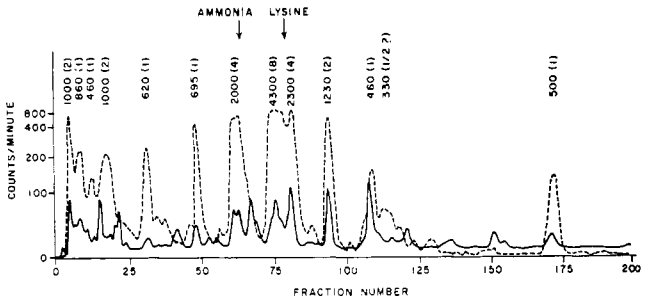

FIGURE 3: Tryptic peptide pattern of protein 9 labeled with [^{3}H]-lysine. Column eluate was split into two equal streams one of which was treated with ninhydrin (—) and the other delivered to a fraction collector for radioisotope measurements (---). Above each radioactive peak is printed the total number of counts it contains with the estimated number of lysine-containing peptides in parentheses.

marized in Table I. These analyses were made on samples that had been hydrolyzed for 18 hr. No attempt has been made to correct the data for the destructive hydrolysis of serine, methionine, and threonine or for the relatively slow liberation of leucine, isoleucine, and valine from peptides. The tryptophan and cysteine content of the proteins has not been measured, but the presence or absence of cysteine was indicated by [^{14}C]iodoacetamide incorporation and qualitative amino acid analysis. In spite of these limitations it is clear that all of the proteins have distinguishable amino acid compositions. More definitive amino acid analyses will be reported in a subsequent paper.

Peptide Mapping. The data in Figure 1 illustrate the precision and reproducibility of the peptide mapping technique of Jones (1964) as applied to the ribosomal proteins. Four independent preparations of the same protein are compared; three are from a B strain of *E. coli* and one is from a K12 strain. The reproducibility of the peptide positions is at worst plus or minus 5 min with the elution time for lysine taken as the standard, and normally the variation is within plus or minus 3 min. Since the total time for development of the Jones column was 450 min, this corresponds to a variation of approximately 1%. The reproducibility of this procedure is underscored by the fact that the peptide maps displayed in Figure 1 were obtained over a 1-year period using several different Dowex 50 columns of the same dimensions.

Another useful feature of the Jones column is illustrated in Figure 1. The third analysis from the top shows several peaks, particularly in the right-hand half of the pattern, which do not appear in the other maps. The preparation of protein used for this analysis was slightly contaminated with a second component as revealed by polyacrylamide gel electrophoresis. Thus, these extra peptides are probably reflections of contamination. In general, the peptide patterns obtained with the Jones column have been quite sensitive to the presence of contaminating proteins. In contrast, the amino acid compositions did not show any significant differences between the contaminated preparation and the three purer ones.

The levels of impurities which can be detected in the purified proteins can be determined experimentally by mixing different proportions of two purified proteins and obtaining a peptide map from the mixture. It is easy to identify the major peptides of a contaminant protein present at 20% of the mixture. However, when there is 10% contamination by a second protein, these peptides are more difficult to detect. Thus, material with less than 5% contamination, according to electrophoretic homogeneity (Hardy *et al.*, 1969), is likely to be adequate for the identification of all of the proteins in the ribosome.

The peptide patterns in Figure 1 are incomplete. The first 50–60 min of column development are not shown. During the first 45 min all of the acidic and neutral amino acids are eluted and only the basic amino acids appear in the same region as the majority of peptides. However, most of the ninhydrin-positive peaks which appear in the first 45 min are not amino acids. Some of them may be acidic peptides. Others seem to be artifacts since they can be obtained from phosphocellulose column fractions which contain no protein.

The peptide patterns for the 21 proteins that were purified according to the procedures of the preceding paper are displayed in Figure 2. Identical patterns have been obtained from at least two independent preparations of each protein; some were obtained as many as six times and most at least three times. One unsatisfactory exception is protein 7 which will be discussed below. Detailed comparisons of the remaining twenty proteins reveal that with the exception of proteins 5 and 9, all have unique peptide maps. This taken together with the amino acid compositions presented in Table I suggests that we have succeeded in purifying at least nineteen unique proteins. Proteins 5 and 9 manifest striking similarities in their peptide maps. Another pair, 2a and 10, also show a vague resemblance to each other.

One way of verifying the homogeneity of the purified proteins is to compare their estimated chemical molecular weight (calculated from their amino acid composition and number of tryptic peptides) with a physical molecular weight measurement. When these two estimates agree it is possible to conclude that the protein is relatively pure. If, however, the chemical estimate is much larger than the physical, the protein must be contaminated.

Chemical Molecular Weights. The chemical molecular weight estimates for all of the proteins described in Figure 2 and Table I are listed in Table II. These calculations were made for each protein by multiplying the number of peptides obtained in a tryptic digest by the average molecular weight of the amino acids and dividing by the mole fractionation of arginine and lysine in each protein. No attempt was made in the calculation to estimate the number of peptides included in the relatively large peaks and extremely small peaks were ignored. Therefore, these estimates of chemical molecular weights must be taken in general as minimum ones. There is reasonable agreement between the molecular weights estimated in this fashion and those obtained from equilibrium sedimentation data for most of the proteins (Table II). In addition, reasonable agreement has been found between such estimates of the molecular weight of both lysozyme and ribonuclease and their actual molecular weights. The data show that the smallest 30S protein is approximately 11,000 daltons.

More accurate chemical molecular weights can be calculated if it is possible to properly weigh those large peaks which contain more than one peptide and if the minor peaks can be included in the count with some confidence. The problem is to determine the number of tryptic peptides even though the intensity of ninhydrin staining is an uncontrollable variable for many peptides. In order to do this we have used

TABLE 1: Amino Acid Compositions.

Amino Acid	Protein																				
	1	2	2a	3	4	4a	5	6	7	8	9	10	11	12	12a	12b	13	14	15	15a	16
Lysine	8.2	4.3	9.5	7.9	5.1	8.3	9.3	7.0	7.6	8.6	10.6	9.5	7.5	8.2	8.8	11.3	11.9	7.4	11.0	11.4	14.7
Histidine	1.5	4.5	0.9	2.2	2.9	2.3	1.9	3.0	2.8	1.4	2.0	1.5	2.6	1.1	1.5	1.8	4.3	5.1	3.0	2.8	3.7
Arginine	5.5	10.6	6.5	6.2	11.6	5.2	8.4	10.3	6.8	11.5	9.9	10.8	6.2	13.5	14.7	12.7	7.1	11.0	11.5	12.9	8.5
Aspartic acid	11.6	12.6	8.1	8.4	9.3	11.3	7.8	9.2	9.1	7.7	7.5	8.3	8.8	6.8	6.1	8.0	8.4	7.9	5.8	5.5	8.9
Threonine	4.3	5.3	4.4	5.0	7.5	5.2	5.1	3.7	4.9	3.6	4.5	3.2	5.6	4.2	6.7	3.2	6.1	6.4	6.2	4.8	3.1
Serine	4.4	3.1	4.6	4.1	2.9	6.2	4.7	6.1	9.4	6.6	4.6	5.4	10.6	5.7	5.3	7.1	5.8	8.4	6.5	7.7	5.8
Glutamic acid	14.1	15.8	11.2	9.1	12.3	10.5	11.0	9.7	10.8	11.9	11.3	12.4	8.9	11.9	9.9	9.5	8.0	11.5	7.4	11.2	9.8
Proline	1.9	3.9	3.8	3.4	4.7	3.2	4.5	3.1	3.2	4.1	4.2	3.3	5.4	2.9	3.1	4.1	5.2	0.9	5.3	4.8	2.2
Glycine	8.9	4.0	8.6	12.5	5.1	8.5	8.8	9.2	12.6	6.3	8.4	8.4	14.0	10.3	6.3	7.1	8.9	7.7	10.4	7.9	4.6
Alanine	9.3	11.0	10.7	12.5	8.0	10.4	10.0	11.9	8.6	11.5	10.3	9.2	8.4	8.5	9.7	10.8	8.1	8.0	7.6	9.7	17.9
Half-cystine	+	−	+	−	−	+	−	−	+	−	−	+	+	−	+	+	+	+	+	+	−
Valine	11.3	6.6	9.3	11.1	8.3	7.0	8.6	8.6	8.3	9.6	8.1	7.1	8.8	7.2	3.9	6.0	7.1	5.3	11.6	7.7	3.4
Methionine	1.0	3.1	3.6	3.2	1.6	2.7	1.7	1.0	1.0	2.0	1.4	0.9	0.5	1.7	0.5	1.6	1.7	0.8	0.4	0.5	2.7
Isoleucine	5.3	4.5	6.5	5.2	8.1	5.1	6.8	6.4	6.3	3.0	6.3	4.1	4.5	5.3	6.7	4.1	4.4	3.1	2.6	2.6	6.4
Leucine	8.2	4.0	7.6	5.7	9.8	8.4	6.5	5.1	4.7	7.6	6.5	10.2	3.6	6.9	6.4	7.5	7.7	11.7	6.6	5.0	4.4
Tyrosine	1.1	2.8	2.1	1.3	0.9	0.8	1.7	1.3	1.6	1.9	2.0	3.7	1.5	3.2	5.9	1.5	1.2	2.1	2.9	2.4	1.4
Phenylalanine	3.1	3.2	2.5	2.4	2.0	4.5	2.5	4.7	2.4	2.6	2.6	2.4	3.1	3.1	3.5	3.1	4.0	2.2	1.4	3.4	2.2
Total	100.0	99.9	99.9	100.0	99.9	99.6	99.3	100.4	100.0	98.8	100.0	100.4	99.9	100.4	99.9	99.6	100.0	99.6	100.1	100.4	99.6
Averaged runs	12	6	8	5	4	9	9	6	4	10	7	4	2	8	4	4	4	7	5	4	5

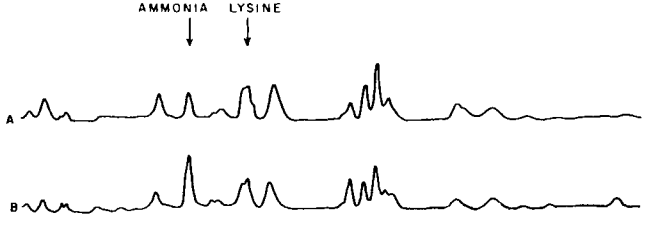

FIGURE 4: Tryptic peptide patterns from protein 2a (A) and aggregate form (B).

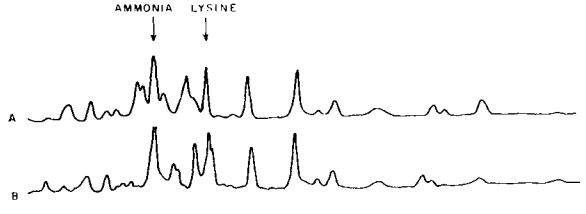

FIGURE 5: Comparison of tryptic peptide patterns of proteins 9 (A) and 5 (B).

TABLE II: Chemical and Physical Molecular Weights.[a]

Protein	Lys and Arg (mole %)	Tryptic Peptides	Mol Wt		[^{3}H]Lys	Lys Mol Wt
			Chemical	Physical		
1	13.7	38	31,000	65,000	36	49,000
2	14.9	24	18,000	18,000		
2a	16.0	27	19,000	17,600		
3	13.8	20	16,200	24,000	17	24,500
4	16.7	22	14,800	16,000		
4a	13.5	33	27,300	30,000		
5	17.7	27	17,000	31,000		
6	17.3	18	11,700	13,500		
7	14.4	20	15,600	10,700		
8	20.1	26	14,500	21,500	15	19,600
9	20.5	26	14,200	33,000	27	28,600
10	20.4	35	19,300	26,700		
11	13.8	20	16,300	18,300		
12	21.6	26	13,500	21,000		
12a	23.5	23	11,000	14,600		
12b	24.0	34	14,200	15,600		
13	19.1	22	13,000	15,000		
14	18.4	26	15,800	13,200		
15	22.5	32	16,000	19,000		
15a	24.3	34	15,700	13,000		
16	23.3	25	12,000	14,000		

[a] The chemical molecular weights were calculated as described in the text using a value of 11,200 daltons/100 moles of amino acids. The physical molecular weights are the average values from Table III. The lysine molecular weights were calculated from [^{3}H]lysine-labeled protein as described in the text.

several different proteins that had been labeled with ^{3}H-labeled lysine. The tryptic digests of the [^{3}H]lysine-labeled proteins were analyzed on a Jones column as described in Methods. Half of the column eluate was collected in a fraction collector for subsequent radioisotope measurements and the other half went through the normal ninhydrin reaction. The data from one such experiment with protein 9 are described in Figure 3; this protein was chosen because of the large discrepancy between its chemical and physical molecular weights. At the extreme right of this peptide map is a major ninhydrin-positive peak which is well separated from other peptides; it contains 500 cpm of ^{3}H-labeled lysine. This peptide can be taken as the unit for the calculation of the number of different peptides in the more complicated regions of the map. Here, we have assumed that 500–700 cpm represents one peptide. On this basis, the number of different lysine-containing peptides in this map is at least 25. This would correspond to a molecular weight of 28,000 daltons for this protein, which is in excellent agreement with the physical molecular weight of 32,000 daltons. This molecular weight is also supported by estimates based on cyanogen bromide digests (see below). Similar experiments were performed with radioactive proteins, 3, 8, and 1 (see Table II). In almost all cases these chemical molecular weight estimates are lower than the physical estimates which gives us confidence that the chemical molecular weights calculated from the unlabeled tryptic peptides do provide a minimum molecular weight estimate.

Physical Molecular Weights. The high-speed sedimentation equilibrium technique of Yphantis (1964) was used for the bulk of our physical molecular weight measurements. The molecular weight of egg white lysozyme was measured in our standard 6 M urea–phosphate buffer. A value of 14,000 daltons was obtained, which is in excellent agreement with the molecular weight of this protein calculated from its sequence by Canfield (1963); this is 14,307 daltons. Generally, our measurements have been reproducible within an error of ±5% but occasional fluctuations of 10% have been observed. The molecular weight measurements obtained by the high-speed procedure are summarized in Table III. The averages for the weight-average molecular weights tabulated here and in Table II are from at least two independently isolated preparations of each protein. In the case of protein 1, some of our early data was omitted from the average for reasons which will be discussed below. We have been unable to detect significant concentration dependence for the apparent molecular weights obtained at the low concentrations that were employed in this study.

In order to be certain that the molecular weight measurements were not distorted by solvent effects or aggregation of the proteins, the measurements of molecular weight for most of the proteins were made in a variety of solvents and in several cases by both low-speed and high-speed procedures. These data are summarized in Table IV. Except for protein 12, no major differences were observed under the following conditions: in 6 M urea–phosphate buffer, in 6 M urea–acetate

TABLE III: Physical Molecular Weights.

Protein	$\bar{v}^a$	Av Mol Wt	Range of Mol Wt	Independent Prepn	Mol Wt Detn
1	0.737	65,000	58,000–72,000	2	5
2	0.716	18,000	17,000–19,000	2	6
2a	0.746	17,600	16,000–18,500	3	4
3	0.739	24,000	22,000–25,000	3	4
4	0.742	16,000	15,000–16,000	2	2
4a	0.735	30,000	27,000–34,000	2	5
5	0.741	31,000	28,000–33,500	3	6
6	0.733	13,500	12,000–15,400	4	7
7	0.727	10,700	8,000–12,000	2	3
8	0.736	21,500	20,000–23,000	4	6
9	0.741	33,000	30,000–35,000	3	4
10	0.738	26,700	25,000–30,000	4	4
11	0.724	18,300	17,000–20,000	2	3
12	0.734	21,000	19,000–22,000	3	3
12a	0.735	14,600	14,000–15,000	2	2
12b	0.736	15,600	14,000–17,100	3	3
13	0.740	15,000	14,000–15,700	2	4
14	0.730	13,200	12,000–14,200	2	6
15	0.739	19,000	19,000–19,100	2	2
15a	0.731	13,000	12,000–13,300	3	3
16	0.736	14,000	13,000–15,100	2	4

[a] The values for $\bar{v}$ were calculated from the amino acid compositions in Table I according to the method of Schachman (1957). All of the data here was obtained by the procedure of Yphantis (1964) and only the data from samples that had negligible deviations from homogeneity are included.

buffer, in 5 M guanidine hydrochloride–phosphate buffer, and after the protein had been alkylated with iodoacetamide. At present we cannot explain the anomalous behavior of protein 12, which has an apparent molecular weight that is dependent on the solvent which is used.

The Identification of the 30S Proteins. Nine of the proteins (4, 6, 8, 12, 12a, 13, 14, 15a, and 16) are unambiguously unique proteins. Their amino acid compositions and tryptic peptide maps are distinctive and reproducible. The chemical molecular weights of these proteins estimated from their amino acid compositions and peptide maps are all in reasonable agreement with their physical molecular weights.

Many more than 21 chromatographic components can be identified among the 30S ribosomal proteins. However, some of these chromatographically separable components are related to each other by their electrophoretic mobilities, amino acid compositions, tryptic peptides, and molecular weights.

At least eight of the 30S proteins have been observed in such multiple chromatographic peaks; these are 2, 2a, 3, 4a, 5, 10, 11, and 12b. Generally in such cases there is both a major chromatographic peak containing the nominal protein as well as a second chromatographic position for a lesser amount of a protein having an identical mobility in polyacrylamide gel electrophoresis. Protein 4a illustrates this well; its major form chromatographs between proteins 4 and 5 and its minor form is a contaminant of protein 3. When the two forms are compared they appear to be almost indistinguishable by all our criteria except chromatographic behavior. Minor differences are seen in the tryptic peptides of the two forms.

The resolution of protein 2 into more than one chromatographic component could be a chromatographic artifact caused by a solvent front. This protein elutes from cellulose phosphate just as the salt gradient begins when chromatography is carried out at pH 6.5. The other examples of multiple forms can be explained in several ways. They could arise from minor chemical modification of the native proteins (deamidation or methionine oxidation) or by the influence of very limited enzymatic modification of the proteins. It is also possible that protein–protein interactions not disrupted by 6 M urea are responsible for these chromatographic anomalies.

The multiple chromatographic forms of proteins 2a and 10 have already been attributed to the formation of disulfide-bonded aggregates by these two proteins (Hardy *et al.*, 1969). The present chemical data demonstrate the validity of this conclusion. In both cases the two forms differ in molecular weight by a factor of two. The amino acid compositions of the dimer and monomer are the same, and as shown in Figure 4, they yield the same peptide maps. Finally, there is at least 1 mole of cysteine/molecule of both protein 2a and 10.

The gross effects of proteolytic degradation of the proteins are seen only after lengthy incubations at room temperature (Hardy *et al.*, 1969). However, there seem to be occasional specific effects on some of the proteins even without lengthy incubations. Although the amino acid composition and tryptic peptide patterns of protein 1 have been quite reproducible, the physical molecular weights of this protein have been more variable than any other protein; values between 17,000 and 72,000 daltons have been obtained. The variation of molec-

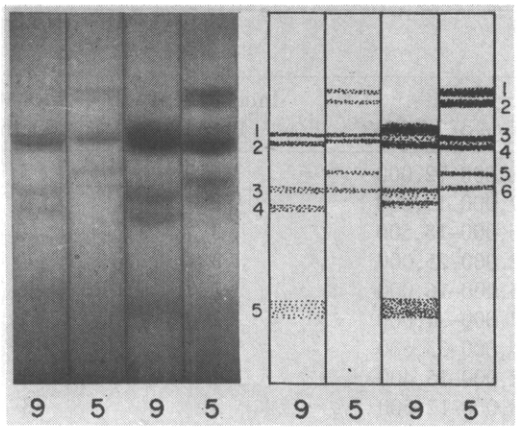

FIGURE 6: Polyacrylamide gel electrophoresis patterns of cyanogen bromide digests of proteins 5 and 9.

ular weight and conservation of chemical composition are precisely what would be expected if this acidic protein were contaminated with proteolytic activity that does not absorb to the cellulose phosphate and appears in the run-off with protein 1. All our recent preparations of protein 1 have apparent molecular weights between 60,000 and 70,000 daltons and the molecular weights are not affected by a broad range of treatments (Table IV).

We have isolated other examples of what are probably enzymatically derived protein fragments. These were identified by their reduced molecular weights and occasional appearance. However, in some cases there was insufficient material to provide definitive identification by peptide analysis.

Protein 7 normally electrophoreses a little faster than protein 6; however, after reduction of the proteins both have the same electrophoretic mobility under our conditions of disc electrophoresis. The successful separation of proteins 6 and 7 allowed us to determine that the peptide maps of protein 6 are much simpler than the mixture of 6 and 7 and the chemical molecular weight of protein 6 is consistent with its physical molecular weight. Although the quality of the peptide maps of three preparations of protein 7 are poor, they are sufficiently different from those of protein 6 to unequivocally distinguish the two proteins, which also have different amino acid compositions. Protein 6 is evidently a unique protein which should be included in the catalogue of 30S components. The classification of protein 7, the smallest and least conspicuous of the purified proteins, is doubtful. We cannot at present make a detailed comparison of the peptide maps of protein 7 and the other proteins. It must therefore be omitted for the time being from the catalogue of 30S proteins because

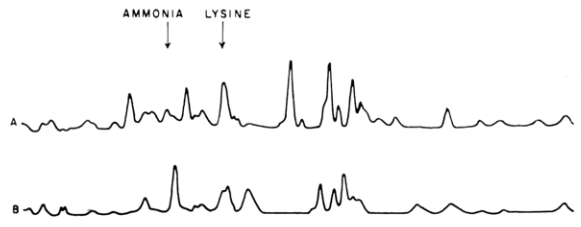

FIGURE 7: Tryptic peptide pattern of protein 10 (A) and protein 2a (B).

TABLE IV: Molecular Weight Measured under Various Conditions.[a]

Protein	Urea–Phosphate	Urea–Acetate	Guanidine–Phosphate	Low Speed
1	61,000		57,000	
2a	18,000	18,000		
	17,000*			
2a aggregate	36,000			36,000
2	18,000	18,000	22,000*	
3	24,000		19,000*	24,000
4a	30,100	27,000	25,000*	
8	22,000	21,000		
	23,000*			
9	33,000			30,000
11	18,300	19,000		
12	21,000	16,000	28,000*	
13	15,000		13,000*	
14	13,200	12,000	13,000*	
15	19,000		18,000*	
15a	13,000		11,000*	
16	14,000	14,000	12,000*	

[a] Urea–phosphate buffer is standard buffer as in Table III; urea–acetate buffer is 6 M urea–0.1 M NaCl–0.01 M sodium acetate–0.01 M β-mercaptoethanol (pH 6.5) and guanidine–phosphate buffer is 5 M guanidine hydrochloride–0.05 M NaH_2PO_4–0.022 M β-mercaptoethanol (pH 6.5). The "low-speed" molecular weights were obtained at ~20,000 rpm as described in methods. An asterisk indicates that the proteins were alkylated with iodoacetamide before dialysis.

there is no way of excluding the possibility that it is a fragment of another protein.

Proteins 5 and 9 have very similar electrophoretic mobilities, peptide maps (Figure 5), amino acid compositions (Table I), and molecular weights (Table III). However, there are significant chemical differences between these two proteins in addition to their distinctive chromatographic behavior. First, the arginine content of the two proteins is distinct and there are tryptic peptides which are present in protein 5 which do not appear in protein 9, and there are peptides which are present only in protein 9. These differences have been consistently observed in four independent preparations of each protein.

The uniqueness of these two proteins is further indicated by the polyacrylamide gel patterns of the peptides obtained by cyanogen bromide digestion (Figure 6). Since such peptide bands are generally diffuse, probably because of their small size, we have included a graphic representation of these peptide patterns along with the photographs of the gels. The methionine contents of the proteins are consistent with the presence of 5 methionines/31,000 daltons for protein 5 and 4 methionines/33,000 daltons for protein 9. From this we would expect that the peptide-bond cleavage adjacent to methionine by cyanogen bromide would produce six peptides from protein 5 and five peptides from protein 9, provided that the methionines are all internal. The data in Figure 5 are con-

sistent with this expectation. Three of the peptides from both proteins are indistinguishable; but protein 5 has three peptides not present in protein 9 and protein 9 has two peptides not present in protein 5. We have recently observed rather marked variation in the relative amounts of proteins 5 and 9 recovered from different ribosome preparations. Furthermore, when the amount of protein 9 is low, we have identified anomalous proteins chromatographing in fractions between proteins 5 and 9. As yet no chemical identification of the proteins has been made; but they are electrophoretically indistinguishable from proteins 5 and 9.

Proteins 2a and 10 (Figure 7) are electrophoretically indistinguishable on polyacrylamide gels and both form dimers in the aggregate state. However, their molecular weights and amino acid compositions are quite different. Similarly there are several regions in the peptide map where they are quite different from one another. For these reasons we suggest that they are unique proteins which contain homologous regions of sequence. However, we have not completely excluded the possibility that protein 2a is a fragment of protein 10.

Protein 15 on two occasions yielded peptide maps consistent with its molecular weight and on two other occasions yielded a peptide map indicating the presence of a second protein. In all four cases the material was electrophoretically homogeneous and the equilibrium sedimentation data revealed no molecular weight heterogeneity.

It is possible to obtain some idea of the tryptic peptide map of the contaminating protein by subtracting peaks in the pattern of pure protein 15 from the pattern obtained with the contaminant present. In this way we obtained a peptide map unlike any that we have displayed in Figure 2; therefore, this contaminant seems to be a unique protein that we have failed as yet to purify.

Discussion

Eighteen of the twenty-one proteins described in the accompanying paper (Hardy *et al.*, 1969) can be identified as unique proteins according to the criteria used here. These are 1, 2, 2a, 3, 4, 4a, 6, 8, 10, 11, 12, 12a, 12b, 13, 14, 15, 15a, and 16. The amino acid composition and peptide maps of each are distinctive. The purity of the proteins is confirmed by the fact that the chemical molecular weights, estimated from amino acid composition and tryptic peptides, are consistent with physical molecular weights, estimated by equilibrium sedimentation analysis. Several possible manifestations of proteolytic activity have obliged us to be cautious in our identifications. Consequently, the unsatisfactory quality of the peptide maps obtained from protein 7 has forced us to defer judgement on its status. The striking similarities between proteins 5 and 9, taken with the variable quantities of protein 9, have suggested that one of these proteins may be an enzymatically derived fragment of the other or that both are fragments of a third protein that we have not recovered. Finally, a protein with apparently unique properties has been identified as an occasional contaminant of protein 15. We conclude that there are probably at least 19 and at most 22 proteins in the 30S ribosomal subunit.

We have observed multiple chromatographic forms for eight of the 30S proteins. Our earlier experience with the ribosomal proteins has obliged us to attribute this phenomenon to the effects of chromatographic artifacts, chemical modifications, and enzymatic modifications of native proteins. However, it is still possible that this "microheterogeneity" is real. If it is real, it could have a profound influence on our notions of the physical homogeneity of the ribosomes.

The existence of so many 30S ribosomal proteins was not anticipated at the outset of this study for the following reasons. The data of previous workers are consistent with a number-average molecular weight for the ribosomal proteins in the neighborhood of 20,000 daltons (Waller, 1964; Möller and Chrambach, 1967) and we have verified this fact for unfractionated 30S protein. Since the 30S particles purified by our procedures contain approximately 250,000 daltons of protein, we expected to purify no more than 12 different proteins from these particles. The identification of approximately twenty proteins whose average molecular weights were known to be much larger than 12,000 daltons before they were purifed was incompatible with the assumption that one copy of every protein is present in every ribosome.

The 16S RNA from the 30S particle has a molecular weight between 530,000 and 560,000 daltons (Kurland, 1960; Midgely, 1965; Stanley and Bock, 1965). The 30S particles that we have employed contain 30–33% protein according to the Lowry and Kjeldahl procedures; this corresponds to between 230,000 and 280,000 daltons of protein per 30S particle. On the other hand, the total mass of protein in a particle that has one copy of each of the nineteen proteins that we have purified would be 410,000 daltons (based on physical molecular weight). Therefore, the sum of molecular weights for 19 30S proteins exceeds by 130,000–180,000 daltons the average mass of protein per 30S particle. Since the reproducibility of our physical molecular weight measurements has been $\pm 10\%$, we believe that this discrepancy is beyond our experimental error. These calculations indicate that the 30S particles may be heterogeneous.

This conclusion is also suggested independently by our preliminary measurements of the relative molar ratios of some of the 30S proteins. An inspection of Figure 2 in the accompanying paper (Hardy *et al.*, 1969) reveals quite striking differences in the mass of each protein recovered from the columns. In several cases this variation is accentuated if the data are corrected for the molecular weights of the proteins. At least one of the proteins (15a) is present in amounts corresponding to one copy for every two or three 30S particles; other proteins are present in similarly small amounts. If these estimates are correct, it follows that every ribosome cannot have the same number of copies of every one of the nineteen proteins so far identified. Therefore, according to two independent criteria the 30S ribosomal subunits prepared by our precedure are heterogeneous.

A possible explanation for these discrepancies is that during the purification of the ribosomes the ammonium sulfate treatment selectively removes some of the ribosomal proteins. Thus the average mass of protein in the particle would be lowered and some of the proteins reduced to minor components. Indeed, we estimate that approximately 80,000 daltons of protein/30S particle are removed in the course of the ammonium sulfate purification (Kurland, 1966; Hardy *et al.*, 1969). This explanation would require that at least some of the proteins removed by ammonium sulfate are identical with the minor proteins that remain in the salt-washed

TABLE V: Comparison of the Molecular Weights Obtained in Two Laboratories.

Geneva Code	Mol Wt	Madison Code	Mol Wt
1	4,500	16	14,000
2a	13,000	13	15,000
3	10,700	6[a]	13,500
4a	9,000	7[a]	10,700
4b	5,500	14	13,200
5	12,800	12	21,000
6	13,700	4	16,000
7	14,500	8	21,500
8a	20,000	3	24,000
10a	5,600	2	18,000
10b	27,600	9	33,000
11	14,000	2a	17,600
13	23,000	1	65,000

[a] The identification of 6 and 7 might be reversed.

particles, and that unwashed ribosomes would contain relatively larger amounts of those proteins which we have identified as minor components. Both of these predictions are contradicted by the data. The proteins removed by ammonium sulfate are different from those remaining with the ribosome. Thus, a number of enzymes are removed (Hardy and Kurland, 1966) and the chromatographic properties of the proteins indicate that they are generally more acidic than the bulk of the remaining ribosomal protein (Hardy *et al.*, 1969). Furthermore, the apparent molar ratios of the basic proteins is not dramatically altered by ammonium sulfate purification (Hardy *et al.*, 1969).

The experiments of Traub *et al.* (1967) lend further support to the conclusion that the minor components are not artifacts. Two of the "split" proteins purified by Traub *et al.* (1967) appear from our data to be minor components; these are B3 and B5, which correspond to our 4 and 12b, respectively. Since both of these minor proteins are absolutely required for the *in vitro* activity of reconstituted ribosomes, they cannot be supernatant contaminants. Therefore, the existence of minor components in the ribosomes cannot be explained away as contaminating supernatant proteins.

We conclude that the apparent deviations from the predicted properties of a homogeneous population of 30S ribosomal subunits may be real. The discrepancies are apparently not due to obvious artifacts but further experimentation is required to prove that ribosomes are heterogeneous *in vivo*.

Some of our results and conclusions are in contrast to those presented by Moore *et al.* (1968). The principal source of disagreement is the molecular weight data, a comparison of which is presented in Table V. The tentative identifications of the proteins purified by Moore *et al.* (1968) with those that we have purified is made on the basis of electrophoretic behavior, amino acid composition, and chromatographic behavior of the proteins in question. There is reasonably good agreement for the molecular weights of seven of the proteins; however, there is a significant deviation in that our molecular weights tend to be slightly larger than those of Moore *et al.* (1968) for those proteins. More striking discrepancies are seen for the molecular weights of the remaining six proteins; in these cases there is as much as a threefold difference in the molecular weight estimates. Since the molecular weight measurements of Moore *et al.* (1968) were done by a procedure which only yields an average molecular weight, there is no information about the physical homogeneity of their samples. It is possible that the group of very small proteins (molecular weight *ca.* 5000 daltons) which were purified by Moore *et al.* (1968) are predominantly fragments of the larger proteins we have purified. It is also conceivable that the protein samples of Moore *et al.* (1968) contained low molecular weight, nonproteinaceous material that was extracted by propionic acid from the material used for sieve chromatography (Stöffler *et al.*, 1967). Whatever the explanation for these discrepancies, it is clear that such differences in the molecular weights must be due to some systematic errors, since in all cases the molecular weights that we have estimated are larger than those of Moore *et al.* (1968). The large number of control experiments, the reproducibility of independent preparations, as well as the chemical molecular weight estimates, make us favor the larger values that we have obtained. Furthermore, in our case, it is doubtful that any proteins with molecular weights as low as 5000 would be recovered after several days dialysis against urea solution. Since the proteins are quantitatively recovered and their relative amounts are unaffected by such dialysis (Hardy *et al.*, 1969), we conclude that such small proteins are not present in our starting material.

A comparison of the molecular weight data for the ribosomal proteins obtained in other laboratories also argues against the existence of a number of small proteins like those isolated by Moore *et al.* (1968); 15 proteins have been purified from the 70S particles of *E. coli* by Kaltschmidt *et al.* (1967); the molecular weights of these proteins are between 9000 and 41,000 daltons. The data of Möller and Chrambach (1967) suggest a weight-average molecular weight of 23,000 daltons which is in excellent agreement with our data but contradicts that of Moore *et al.* (1968) who found a weight-average molecular weight of 15,000 daltons.

The possibility that the 30S ribosomes of *E. coli* are heterogeneous has not been definitively proven by our data. However, the evidence is sufficiently suggestive that some speculation about the significance of our conclusions is required, if only to provide models that can be tested by experiment. Traub *et al.* (1968) have demonstrated that in at least two cases the ribosomes are completely inactive if a single protein is omitted during reconstitution of the ribosomes. This strongly suggests that there is one copy of each of these proteins in every functional ribosome, but it does not exclude the possibility that other proteins are present in some of the ribosomes and absent from others. There is the possibility that the ribosome can exist *in vivo* in several functionally distinct states: resting, initiating a polypeptide chain, translating a message, and chain terminating. There could be some ribosomal proteins that are associated with or disassociated from the ribosome at each of these stages. Thus, we envision a basic core of ribosomal proteins that are present in all of the particles, and a second group of perhaps five or six proteins that are exchanged from one 30S particle to the other. The presence or absence of these proteins would determine the functional state of the 30S ribosomal subunit.

We favor this model because it does not complicate our ideas about the fundamental nature of ribosome function, as would the existence of more than one distinct kind of ribosome. We are suggesting that when the specific steps in protein synthesis are analyzed in more detail, it may turn out that the operationally defined ribosomal particle functioning at each stage is slightly different from that at every other stage in protein synthesis.

Acknowledgments

We are indebted to M. Nomura and O. Smithies for their advice and to R. Erickson, K. Mellenberger, and L. Wollangk for technical assistance.

References

Canfield, R. E. (1963), *J. Biol. Chem. 238*, 2698.
Hardy, S. J. S., and Kurland, C. G. (1966), *Biochemistry 5*, 3676.
Hardy, S. J. S., Kurland, C. G., Voynow, P., and Mora, G. (1969), *Biochemistry 8*, 2897 (this issue; preceding paper).
Hill, R. L., and Delaney, R. (1967), *Methods Enzymol. 11*, 339.
Jones, R. T. (1964), *Cold Spring Harbor Symp. Quant. Biol. 29*, 297.
Kaltschmidt, E., Dzionara, M., Donner, D., and Wittman, H. G. (1967), *Mol. Gen. Genetics 100*, 364.
Kawahara, K., and Tanford, C. (1966), *J. Biol. Chem. 241*, 3228.
Kurland, C. G. (1960), *J. Mol. Biol. 2*, 83.
Kurland, C. G. (1966), *J. Mol. Biol. 18*, 90.
Midgely, J. E. M. (1965), *Biochim. Biophys. Acta 108*, 340.
Möller, W., and Chrambach, A. (1967), *J. Mol. Biol. 23*, 377.
Moore, P. B., Traut, R. R., Noller, H., Pearson, P., and Delius, H. (1968), *J. Mol. Biol. 31*, 441.
Richards, E. G., Teller, D. C., and Schachman, H. K. (1968), *Biochemistry 7*, 1054.
Schachman, H. K. (1957), *Methods Enzymol. 4*, 32.
Spackman, D. H., Moore, S. and Stein, W. H. (1958), *Anal. Chem. 30*, 1190.
Stanley, W. M., and Bock, R. M. (1965), *Biochemistry 4*, 1302.
Steers, E., Jr., Craven, G. R., Anfinsen, C. B., and Bethune, J. L. (1965), *J. Biol. Chem. 240*, 2478.
Stöffler, G., Rudloff, V., and Wittman, H. G. (1967), *Mol. Gen. Genetics 101*, 70.
Traub, P., Hosokawa, K., Craven, G. R., and Nomura, M. (1967), *Proc. Natl. Acad. Sci. U. S. 58*, 2430.
Waller, J. P. (1964), *J. Mol. Biol. 10*, 315.
Yphantis, D. A. (1964), *Biochemistry 3*, 297.