Rodbard, D., & Feldman, H. A. (1975) *Methods Enzymol. 36*, 3–16.
Scatchard, G. (1949) *Ann. N.Y. Acad. Sci. 51*, 660–672.
Smith, R. M., & Martell, A. E. (1976) *Critical Stability Constants*, Vol. 4, Plenum Press, New York.
Solomon, E. I., Eickman, N. C., Himmelwright, R. S., Hwang, Y. T., Plon, S. E., & Wilcox, D. E. (1982) in *Physiology and Biochemistry of Horseshoe Crabs* (Bonaventura, J., Bonaventura, C., & Tesh, S., Eds.) pp 189–230, A. R. Liss, Inc., New York.
Stumm, W., & Morgan, J. J. (1981) in *Aquatic Chemistry*, Chapter 3, Wiley-Interscience, New York.
Turner, B. W., Pettigrew, D. W., & Ackers, G. K. (1981) *Methods Enzymol. 76*, 596–628.
van Driel, R., & von Bruggen, E. F. J. (1974) *Biochemistry 13*, 4079–4083.
Van Holde, K. E., & van Bruggen, E. F. J. (1971) *Biol. Macromol. 5 (Part A)*, 1–55.
Van Holde, K. E., & Miller, K. I. (1982) *Q. Rev. Biophys.*, 1–129.
Whaling, P. J., Barber, R. T., & Paul, J. C. (1977) *Rep.—Water Resour. Res. Inst. Univ. N.C. 123.*
Wright, D. A. (1977a) *J. Exp. Biol. 67*, 137–146.
Wright, D. A. (1977b) *J. Exp. Biol. 67*, 147–161.
Wright, D. A., & Brewer, G. C. (1979) *Mar. Biol. (Berlin) 50*, 151–156.
Zolla, L., Kuiper, H. A., Vecchini, P., Antonini, E., Brunori, M., & Wyman, J. (1978) *Eur. J. Biochem. 87*, 467–473.

Interaction of R17 Coat Protein with Synthetic Variants of Its Ribonucleic Acid Binding Site†

Jannette Carey, Peggy T. Lowary, and Olke C. Uhlenbeck*

ABSTRACT: The specificity of the interaction between R17 coat protein and its site of translational repression on R17 RNA was studied by enzymatically synthesizing 23 sequence variants of the RNA binding site and measuring their affinity to the coat protein by a nitrocellulose filter binding assay. Experiments using oligomers truncated on the 3′ and 5′ termini allowed precise determination of the edges of the binding domain. Several oligomers which disrupted one or more of the base pairs in the binding site failed to bind coat protein, establishing the importance of RNA secondary structure for the interaction. Substitution at two single-stranded positions with each of the common bases affected K_a very differently. In one case, K_a was reduced substantially no matter which base was substituted for an adenine. At the other position, when a uracil was substituted with a purine, K_a decreased 10–100-fold, whereas when it was substituted by a cytosine, K_a increased about 5-fold. These studies indicate that the protein and the RNA hairpin loop interact over an extensive area and that several different types of contacts form to stabilize the complex.

The translational repression of the bacteriophage R17 replicase gene by the phage coat protein is a convenient system in which to study the molecular basis of a specific protein–RNA interaction. About 10 min after phage infection, phage coat protein accumulates to a concentration high enough to saturate a specific binding site on the phage RNA (Lodish & Zinder, 1966; Bernardi & Spahr, 1972). Since this site contains the initiator region for translation of the phage replicase subunit, repression of synthesis of replicase protein occurs. The binding site for coat protein consists of a single small RNA hairpin (Steitz, 1974; Jansone et al., 1979), the total synthesis of which has been recently achieved (Krug et al., 1982). The synthetic fragment binds coat protein with unit stoichiometry and an equilibrium constant identical with that of intact R17 RNA (Carey et al., 1983). Since other RNAs compete poorly with the fragment for coat protein binding, the interaction appears to be highly specific.

In this paper, we report the synthesis and coat protein binding properties of 23 sequence variants of the RNA binding site. The approach of using synthetic variants to define a protein binding site has been used successfully to identify the type and number of protein–nucleic acid contacts in the *lac* repressor–operator interaction (Caruthers, 1980). Since several of the physical properties of the R17 system differ substantially from those of *lac* (Carey & Uhlenbeck, 1983), it was of interest to compare the binding properties of variants. The results show that, like the *lac* system, there are several points on the RNA which contact the protein and other sites that are not essential for binding. In addition, the RNA secondary structure is found to be an important element in the specific interaction with the protein.

Materials and Methods

Synthesis of Variant Binding Fragments. RNA binding fragments were synthesized enzymatically according to the general protocol described by Krug et al. (1982) and summarized in Figure 1. The 21-nucleotide fragment is made by the ligation of a dodecanucleotide acceptor with a nonanucleotide donor. These two "half"-molecules were each prepared by the addition of two "quarter"-molecules. The synthesis of the 23 variants is summarized in Table I. In most cases, a variant was prepared by introducing one or sometimes two unique quarter-molecules in the synthetic scheme. These are identified by underlining in Table I.

Variant quarter-molecules were prepared in polynucleotide phosphorylase reactions using buffer A [0.4 M NaCl, 10 mM $MgCl_2$, and 0.2 M tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), pH 8.2] or in RNA ligase reactions with buffer B [20 mM $MgCl_2$, 3 mM dithiothreitol, and 50 mM *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid (Hepes), pH 8.0]. Reactions were terminated by heating to 65 °C for 5 min. In many cases, 100 μg/mL bacterial alkaline

† From the Department of Biochemistry, University of Illinois, Urbana, Illinois 61801. *Received April 20, 1983.* This work was supported by a grant from the National Institutes of Health (GM 19059).

0006-2960/83/0422-4723$01.50/0

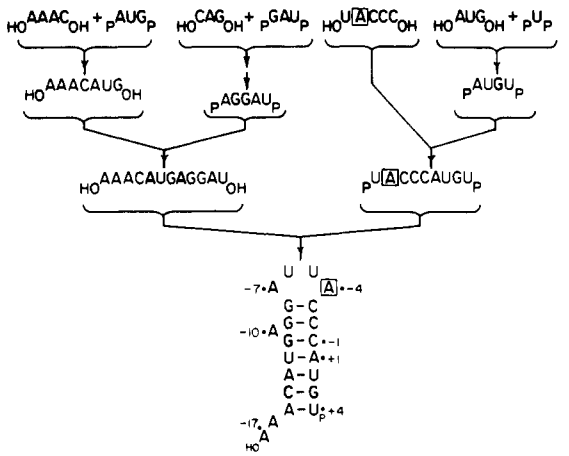

FIGURE 1: Synthesis of the R17 coat protein binding site adapted from Krug et al. (1982). Residues are numbered with respect to the first nucleotide of the synthetase gene. Changing the boxed nucleotide in UAC_3 results in a variant at −4.

Table I: Synthesis and Numbering of Variant Binding Fragments[a]

natural sequence	
[AAACAUG + AGGAU] + [UACCC + AUGU]	1
5′ truncates	
[AACAUG + AGGAU] + [UACCCAUGU]	2
[AUG + AGGAU] + [UACCCAUGU]	3
[AGGAU] + [UACCCAUGU]	4
3′ truncates	
[AAACAUGAGGAU] + [UACCC + AUG]	5
[AAACAUGAGGAU] + [UACCC + AU]	6
[AAACAUGAGGAU] + [UACCC + A]	7
[AAACAUGAGGAU] + [UACCC]	8
[AAACAUGAGGAU] + [UACC]	9
−4,−5 variants	
[AAACAUGAGGAU] + [XYCCC + AUGU]	10-16[b]
5′ stem	
[AAACAUG + AAAAU] + [UACCCAUGU]	17
[AAACAUG + GGAU] + [UACCCAUGU]	18
3′ stem	
[AAACAUGAGGAU] + [UACC + AUGU]	19
[AAACAUGAGGAU] + [UACCCC + AUGU]	20
[AAACAUGAGGAU] + [UAAAA + AUGU]	21
[AAACAUGAGGAU] + [UAUUC + AUGU]	22
[AAACAUGAGGAU] + [UACCU + AUGU]	23
[AAACAUG + AAAAU] + [UAUUU + AUGU]	24

[a] Unique quarter-molecules are underlined. Brackets indicate the two half-molecules used in the final ligation. [b] Group of seven variants where XY is replaced by UG, UC, UU, GA, CA, AA, and AC, respectively.

phosphatase was added and incubation continued for an additional 3 h at 37 °C to remove terminal phosphates from oligomers or to degrade nucleoside di- or triphosphates that interfered with product purification. Oligomers were purified by descending paper chromatography as described by Krug et al. (1982) using solvents, A, B, C, or D, which correspond to 60:40, 50:50, 30:70, or 70:30 v/v ratios of 95% ethanol and 1 M ammonium acetate. The syntheses of quarter-molecules not previously described in Krug et al. (1982) are reported here.

ApApCpApUpG was prepared in a 0.45-mL reaction by incubating 1 mM ApApC, 1 mM pApUpGp, 1.3 mM ATP, and 75 μg/mL RNA ligase in buffer B for 12 h at 12 °C. After treatment with alkaline phosphatase, the product was purified in 88% yield by using solvent B.

The seven pentamers XpYpCpCpC were prepared in 0.5-mL reactions by incubating 6 mM XpY, 20 mM CDP, and 10 units/mL polynucleotide phosphorylase in buffer A for 36 h at 37 °C. After treatment with phosphatase, the oligomers were purified by using solvent A. UpApCpC and UpApCpCpCpC were obtained as side products of the UpApCpCpC synthesis of Krug et al. (1982). UpApUpU and UpApUpUpU were prepared from UpA and UDP in an identical fashion except that solvent D was used in purification. UpApApApA was prepared from UpA and ADP in a similar fashion except that incubation was for 18 h and solvent C was used in purification. In each of these equilibrium reactions, the yields were about 15% of the dimer incorporated into product.

UpApCpCpU was prepared in a 1.0-mL reaction containing 1.2 mM UpApCpC, 1.5 mM pUp, 2.5 mM ATP, and 195 μg/mL RNA ligase in buffer B. After 12 h at 12 °C followed by phosphatase treatment, the product was purified in 75% yield by using solvent D. UpApUpUpC was prepared in a 0.20-mL reaction by incubating 0.5 mM UpApUpU, 1.35 mM pCp, 2.5 mM ATP, and 245 μg/mL RNA ligase for 16 h at 12 °C. Following phosphatase treatment, the product was purified by using solvent D in 39% yield. ApApApApU was prepared by RNase A digestion of poly(A-U) as described in Uhlenbeck & Cameron (1977).

pGpGpApUp was prepared in three steps. First, ApApCpGpGpApUp was made by incubating 0.5 mM ApApCpG, 0.5 mM pGpApUp, 1 mM ATP, and 80 μg/mL RNA ligase in buffer B for 18 h at 12 °C. The heptamer was purified with solvent B in 60% yield. Ribonuclease A digestion of this product and 5′-phosphorylation of the resulting GpGpApUp were accomplished simultaneously in a reaction containing 0.24 mM ApApCpGpGpApUp, 2.5 mM ATP, 17 μg/mL RNase A, and 13 units/mL pseT 1 polynucleotide kinase. Incubation at 37 °C for 2 h followed by chromatography in solvent B resulted in a 27% yield of pGpGpApUp.

The synthesis of variant half-molecules from quarter-molecules followed the general protocol of Krug et al. (1982). For the synthesis of 3′ half-molecules, 0.1-mL reactions contained 0.8 mM donor, 0.8 mM acceptor, 2 mM ATP, and 200 μg/mL RNA ligase in buffer B. After incubation for 8 h at 25 °C, the product was purified by paper chromatography using solvent A. Yields ranged from 10 to 70%. For the synthesis of 5′ half-molecules, the 0.1-mL reactions contained 0.1 mM acceptor, 60 μM donor, 2.5 mM ATP, and 250 μg/mL RNA ligase in buffer B. After incubation for 8–12 h at 12 °C, alkaline phosphatase was added. The product was purified by using solvent B. Yields ranged from 50 to 70%.

The binding fragment and the 23 variants were prepared with an internal ^{32}P label from half-molecules in two steps. In the first step, the 3′ half-molecule was 5′-^{32}P labeled in a 10-μL reaction containing 20–50 μM oligomer, 30 μM [γ-^{32}P]ATP (160 Ci/mmol), and 300 units/mL pseT 1 polynucleotide kinase in buffer B. After incubation for 2 h at 37 °C, the reaction was heated for 5 min at 65 °C to inactivate the enzyme. In the second step, the 5′ half-molecule acceptor was added to the 5′-^{32}P-labeled donor in a 10-μL reaction containing 3 μL of the above kinase reaction, 14 μM 5′ half-molecule, 1 mM ATP, and 300 μg/mL RNA ligase in buffer B. After incubation for 12 h at 14 °C, the reaction was terminated by the addition of 10 μL of 0.6% bromophenol blue and 0.3% xylene cyanol in 10 M urea, and each reaction was applied in two 1-cm-wide wells of a 150 × 260 × 0.75 mm gel of 20% polyacrylamide–7 M urea. Electrophoresis was at 1000 V for 6 h. Product oligomers were localized by autoradiography, excised, and eluted from the crushed gel slice by soaking at 4 °C for 12 h in 1.5 mL of 0.1 M Tris-HCl, pH 7.5, 0.1 mM ethylenediaminetetraacetic acid (EDTA), and 200 μg/mL *Escherichia coli* mixed tRNA. The eluate was filtered and ethanol precipitated. Recoveries ranged from 0.5

to 5 μCi. The purified RNAs were dissolved in water at about 5 μCi/mL and stored at −20 °C. No degradation of the purified RNAs was observed on analytical gels in up to 60 days of storage.

Product Characterization. Digestion of internally ^{32}P-labeled variants with mixed nucleases to give nucleoside 3′-monophosphates or with snake venom phosphodiesterase to give nucleoside 5′-monophosphates and analysis by two-dimensional thin-layer chromatography are described in Krug et al. (1982). Ribonuclease T_1 digestion of 1–10 nCi of each variant was performed in 15-μL reactions containing 20 mM sodium citrate, pH 5.3, 1 mM EDTA, and 7 M urea. Reactions were incubated 15 min at 50 °C, and then 0.5 unit of RNase T_1 was added in 1 μL and incubation continued at 50 °C for an additional 15 min. Ribonuclease A digestions were performed on 1–10 nCi of each variant in 5-μL reactions containing 40 mM Hepes, pH 7.5, and 0.6 ng of RNase A. Incubation was for 2 h at 37 °C. Partial base hydrolysis of 5′-^{32}P-labeled 21-mer was carried out as described in Carey et al. (1983). When base hydrolysis and RNase A or T_1 digestions were complete, an equal volume of 10 M urea–dye mixture was added, and the reactions were applied to analytical polyacrylamide gels.

Nitrocellulose Filter Assay. The filter retention assay for the interaction of R17 coat protein and its binding fragment is described in detail in Carey et al. (1983). Since each radiolabeled variant is available only in relatively small molar amounts, the association constant is determined from a protein excess binding experiment. In a series of reactions, a constant, low concentration of labeled RNA (~10 pM) in TMK buffer (10 mM magnesium acetate, 80 mM KCl, and 0.1 M Tris-HCl, pH 8.5) is mixed with varying concentrations of coat protein (10^{-10}–10^{-6} M). After incubation at 2 °C for 20 min, an aliquot of each reaction is filtered without dilution or subsequent washing on a nitrocellulose filter. A binding curve is plotted as the fraction of input RNA bound as a function of coat protein concentration. Binding curves typically reach a plateau at which less than 100% of the input RNA is bound. The fraction of RNA bound at the plateau is assumed to represent complete binding of active RNA molecules (Carey et al., 1983). The coat protein concentration required to reach half-plateau binding of the RNA is equal to the inverse of the association constant (K_a).

Results

Synthesis of Variant RNAs. The 23 variant coat protein binding fragments compiled in Table I were synthesized enzymatically according to the general scheme of Krug et al. (1982). As indicated by the boxed nucleotide in Figure 1, this scheme permits the synthesis of a variant fragment by changing just one of the starting oligomers. Since only radiochemical amounts of each variant were needed to determine the binding constant to the protein, only 1–10 nmol of each variant half-molecule was made. The ^{32}P label was introduced during the final joining of the two half-molecules, thereby avoiding the additional steps necessary for subsequent labeling of purified variants. The products of the final ligation reaction were applied to a denaturing polyacrylamide gel for purification of each variant. An autoradiogram of the preparative gel for the synthesis of variants **19**, **20**, and **21** is shown in Figure 2A. The major bands have the expected relative mobilities for oligomers 20, 22, and 21 nucleotides long, respectively. The resolution of the gel permitted excision of the major band free from side products one nucleotide shorter such as those seen for variants **19** and **21** in Figure 2A. Each variant was eluted from the gel slice, ethanol precipitated, and

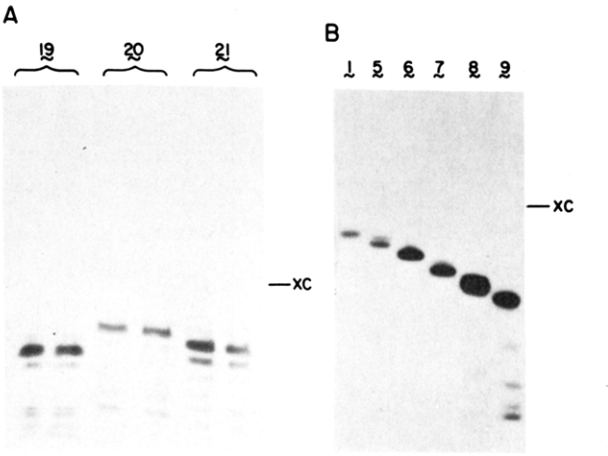

FIGURE 2: Autoradiograms of preparative (A) or analytical (B) polyacrylamide gels of the indicated variants. XC is xylene cyanol marker.

reanalyzed on an analytical gel. An example of an analytical gel of variants **5–9** is shown in Figure 2B. In all cases, a major band of the correct length was recovered, although some variants contained contaminants which constituted less than 10% of the radioactivity of the main band.

The sequence of **1** synthesized previously (Krug et al., 1982) was verified by enzymatic RNA sequence determination. The molecules prepared in this work, however, were labeled internally with the [^{32}P]phosphate between positions −5 and −6 and therefore could not be sequenced by those methods. Different criteria for their identity and homogeneity had to be applied. Since each variant was prepared by combining a well-characterized new quarter-molecule with the previously characterized quarter- and half-molecules, the sequence verification focused on the accuracy of the final ligation reaction. As shown for nine variants in Figure 3A, mixed nuclease digestion of each variant to nucleoside 3′-monophosphates gave [^{32}P]UMP. This result was expected since each 5′ half-molecule had a 3′-terminal uridine residue. In addition, snake venom phosphodiesterase hydrolysis of each variant to nucleoside 5′-monophosphates gave the ^{32}P-labeled mononucleotide corresponding to the 5′-terminal nucleotide of the donor molecule, as shown for nine variants in Figure 3B. Thus, the expected UpN internucleotide linkage between donor and acceptor was formed in each case.

An additional criterion for the identity of a variant is that a labeled oligomer of the correct size results from total RNase A or RNase T_1 hydrolysis of the variant. The products of the digestion reactions are separated on a 20% or 30% polyacrylamide–urea gel and their lengths compared with those of the hydrolysate of **1**. As can be seen in Figure 4A, digestion with RNase A gives a single labeled product of the same size for each of nine representative variants. Comparision of this product with the base hydrolysate and size markers indicates that it is a hexamer, as expected on the basis of the sequences of all the variants in Figure 4A. For **1**, this hexamer has the sequence GAGGAUp. Figure 4B shows that RNase T_1 digestion of seven variants yields unique labeled products of different sizes. For **1**, this product is a decamer with the sequence AUUACCCAUGp. The products resulting from RNase T_1 digestion of the remaining variants in Figure 4B have the mobilities expected on the basis of their sequences.

The purified variants are thus of the expected lengths, and they are homogeneous; they give products of the predicted sizes upon digestion with either RNase A or RNase T_1, and the ligase joints have the correct identity and orientation of the

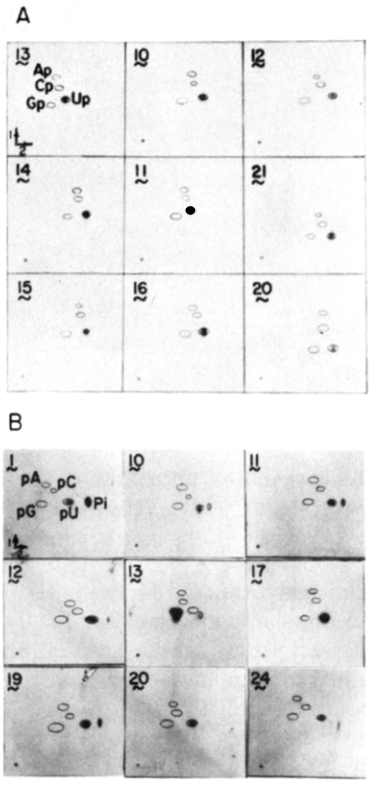

FIGURE 3: Autoradiograms of two-dimensional thin-layer chromatograms identifying the nucleotides flanking the [^{32}P]phosphate in the indicated variants. (A) Digestion to nucleoside 3′-monophosphates with mixed nucleases. (B) Digestion to nucleoside 5′-monophosphates with snake venom phosphodiesterase.

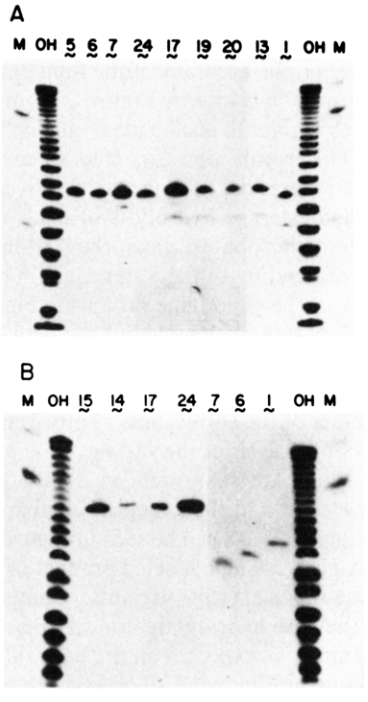

FIGURE 4: Autoradiograms of gels analyzing the products of ribonuclease A (A) or ribonuclease T_1 (B) digests of the indicated variants. The lanes marked OH contain a partial alkaline digest of 5′-^{32}P-labeled 21-mer, and those marked M are a mixture of standards 21, 15, and 4 nucleotides in length.

adjoining bases. Thus, in each synthesis, the desired product was recovered and shown to have the characteristics predicted on the basis of its presumptive sequence.

Coat Protein Binding to Truncated Variants. The results of coat protein binding experiments with **1**, the 21-nucleotide fragment synthesized by Krug et al. (1982), and with five variants truncated on the 3′ terminus are compared in Figure 5A. As successive nucleotides are removed from the 3′ terminus of the 21-mer, K_a at first does not change and then drops abruptly, indicating the edge of the binding domain. Thus, variants **5** and **6**, which are missing one and two 3′-terminal nucleotides, bind the protein just as well as **1**. Variant **7**, which terminates at A_{+1}, has a 4-fold lower K_a than **1**, and variant **8**, which terminates at C_{-1}, has an estimated K_a 50-fold lower than **7**. Little binding is detectable for **9**, which terminates at C_{-2}. It is unclear exactly which nucleotide constitutes the 3′ terminus of the binding domain. The relatively small difference in K_a between variants **6** and **7** suggests that U_{+2} may contact the protein weakly. Alternatively, U_{+2} may be required to help position nearby nucleotides for proper protein binding. The low K_a for variant **8** clearly implicates A_{+1} as part of the protein binding site. Since **8** terminates with a 3′-hydroxyl and **7** with a 3′-phosphate, the 50-fold difference in K_a between **7** and **8** is the result of removing two phosphates and nucleoside A_{+1}.

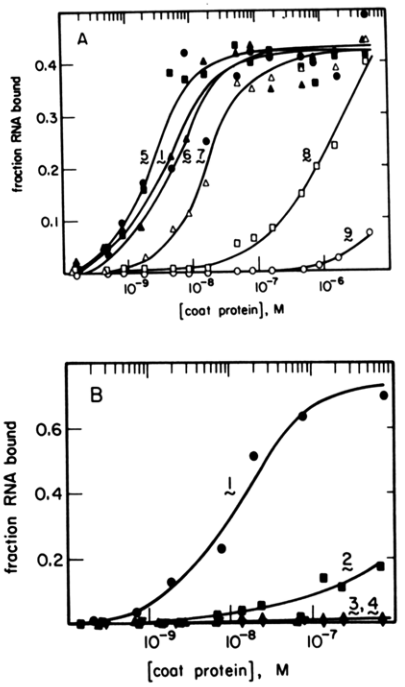

FIGURE 5: Coat protein binding to variants truncated at the 3′ terminus (A) or the 5′ terminus (B) in TMK buffer at 2 °C. K_a values which can be reliably determined are 3×10^8 (**1**), 4×10^8 (**5**), 2×10^8 (**6**), and 5×10^7 (**7**) M^{-1}.

Three nucleotides can thus be removed from the 3′ terminus of the 21-mer without substantially altering its affinity for the protein. This result clearly shows that the base pairs at the bottom of the hairpin stem are dispensable. The much lower K_a for the variant missing A_{+1} may be the result of deleting an essential contact between the protein and the A_{+1}·U_{-12} base pair. Alternatively, the A_{+1}·U_{-12} base pair may be essential for maintaining the correct structure in the upper part of the stem. Although a hairpin loop with three G·C base pairs and a bulged nucleotide would be expected to be stable at 2 °C, extensive fraying of the terminal base pair could alter the environment of the potentially important A_{-10}. The presence of the A_{+1}·U_{-12} base pair would reduce such fraying and maintain proper orientation of the upper part of the hairpin loop.

Coat protein binding to a series of variants truncated on the 5′ terminus is shown in Figure 5B. Although the series is less complete, a similar picture emerges. In this case, variant **2**,

which lacks only A_{-17}, already binds with a greatly reduced K_a, indicating that this nucleotide is essential. Variants **3** and **4**, which lack four and seven nucleotides from the 5′ terminus, respectively, show no detectable binding. Since extending the sequence of **1** further in the 5′ direction does not increase K_a (Carey et al., 1983), the edge of the binding site is at A_{-17}. Since A_{-17} is in a single-stranded portion of the molecule, it is most likely that A_{-17} directly contacts the protein and thereby contributes to K_a. The interaction of A_{-17} with the protein appears to be insensitive to the secondary structure of the RNA. Since residues G_{+3} and U_{+4} can be removed without affecting K_a, residues C_{-14} and A_{-15} do not need to be base paired for proper interaction of A_{-17} with the protein. In addition, elongating the 21-mer with two U residues on the 3′ end, forming two additional base pairs, does not lead to altered coat protein binding (J. Carey, V. Cameron, M. Krug, and P. de Haseth, unpublished results). Thus, A_{-17} can be base paired and still interact normally with the protein. These data suggest that the contact between A_{-17} and the coat protein is more likely to involve the sugar–phosphate backbone than the base functional groups. Variants with sequence changes at position −17 will be required to establish this point.

Variants at Positions −4 and −5. The coat protein binding properties of variants in which the adenine at position −4 is replaced by each of the other common bases are shown in Figure 6A. In each case, very poor binding of variant RNAs is observed. Even at the highest protein concentration used (5 μM), the binding curves do not reach saturation. Although it would be possible to estimate values of K_a for these variants by extrapolation, at least two limitations of the assay make it unlikely that such values would be accurate. First, when the input coat protein concentration in the assay exceeds 100 nM, uncomplexed protein bound to the filter is capable of binding free RNA passing through the filter (Carey et al., 1983). This effect could lead to an erroneously high estimate of K_a for variants which bind poorly. Second, coat protein is known to aggregate at concentrations above 10 nM (Shafranski et al., 1975). Since the RNA binding properties of the aggregates are unknown, it is difficult to evaluate the filter-bound radioactivity at high protein concentrations.

A third potential difficulty in evaluating weak K_a values with the filter binding assay is that the complex might undergo significant dissociation during the filtration time. Since the half-life of the complex with **1** is about 1 min and the filtration time is a few seconds, variants with K_a values 100-fold weaker than **1** might be expected to dissociate during filtration. This possibility was tested by obtaining a binding curve for variant **15** with or without two 1-mL washes of binding buffer. Just as was previously found for **1** (Carey et al., 1983), washing the filter did not alter the binding curve. It is thus unlikely that dissociation of complexes on the filter alters the K_a observed for variants.

Although it is not possible to deduce the K_a values for variants **10**, **11**, and **12** from the data in Figure 6A, it is clear that a substantial reduction in K_a occurs when the adenine is changed to any other base. These data therefore clearly identify A_{-4} as a site of contact with the protein. The interaction appears to be highly specific at this site since neither another purine nor bases with some of the same functional groups can even partially replace the adenine.

As shown in Figure 6B, the result of making single base substitutions at position −5 in the hairpin loop is somewhat different from the effects seen at the adjacent position. The magnitude of the effect on coat protein binding of a base change at position −5 depends on the identity of the replacement base. Changing U_{-5} to an adenine results in only about a 10-fold decrease in K_a, while changing it to a guanine results in about a 100-fold decrease. These data suggest that the base substitutions not only prevent formation of one or more contacts between the coat protein and the uracil at −5 but also may introduce unfavorable constraints that prevent additional contacts as well. The adenine at this position is in some way less disruptive than the guanine.

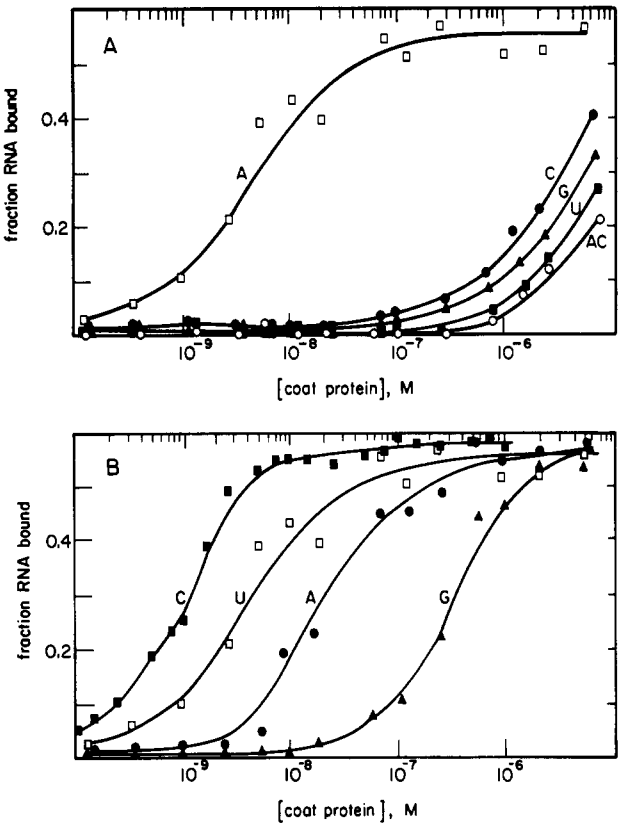

FIGURE 6: Coat protein binding to variants at positions −4 and −5 in TMK buffer at 2 °C. (A) Variants **1** (□), **10** (▲), **11** (●), and **12** (■) which contain the indicated nucleotide at position −4 and variant **16** (○) containing the change at positions −4 and −5. (B) Variants **1** (□), **13** (▲), **14** (■), and **15** (●) which contain the indicated nucleotide at position −5. K_a values which can be reliably determined are 3×10^8 (**1**), 4×10^6 (**13**), 1.5×10^9 (**14**), and 4×10^7 (**15**) M^{-1}.

The substitution of U_{-5} with a cytosine results in a variant which binds slightly better than **1** to coat protein. Although the data are somewhat variable, several determinations with the same set of protein dilutions give a K_a for the C variant about 5-fold higher than the natural sequence. It is therefore clear that the contact between coat protein and the RNA at position −5 must involve structural features common to both pyrimidines. The reason the variant containing cytosine actually binds somewhat better than that containing uracil is presently unclear.

Variant **16** (Figure 6A) is a double substitution in which positions $U_{-5}A_{-4}$ are changed to $A_{-5}C_{-4}$. This variant does not bind coat protein to an appreciable extent. Since the U → A substitution has only a small effect on K_a, it is presumably the A → C substitution that is primarily responsible for the poor binding.

Other Variants. The three base pairs which comprise the upper part of the helix appear to be essential for coat protein binding. Variant **17**, in which G_{-8} and G_{-9} are replaced with A residues, and variant **21**, which substitutes C_{-1}, C_{-2}, and C_{-3} with A residues, do not bind coat protein to a detectable extent (Figure 7). Variant **19**, which is missing one of the C residues and therefore cannot make all three G·C pairs, also fails to

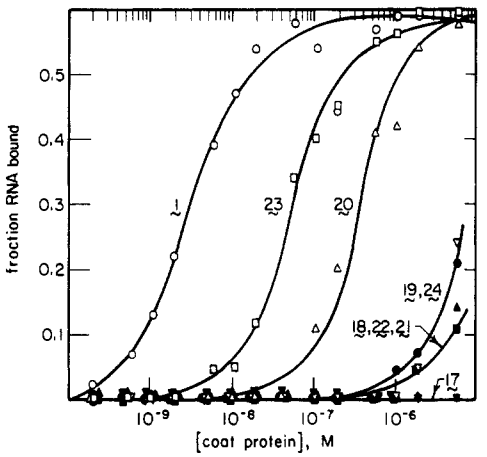

FIGURE 7: Coat protein binding to stem variants in TMK buffer at 2 °C. K_a values which can be reliably determined are 3×10^8 (**1**), 3×10^6 (**20**), and 2×10^7 (**23**) M^{-1}.

bind coat protein. Since these three variants disrupt the secondary structure in addition to changing the nucleotide sequence of the binding site, it is not possible to determine whether the three G·C base pairs directly contact the protein or are needed simply to maintain the orientation of essential single-stranded residues. In an attempt to distinguish between these possibilities, two variants were prepared which were expected to maintain the secondary structure but change the sequence. Variant **22** replaces C_{-2} and C_{-3} by U residues and should thus permit two adjacent G·U pairs to form in the upper part of the helix. The finding that **22** does not bind coat protein suggests that an essential contact in the helical region has been lost. However, the low stability of adjacent G·U pairs and the possibility that **22** can adopt two alternate secondary structures prevent this from being a firm conclusion.

Variant **24** was constructed to determine the effect on coat protein binding of replacing the G·C base pairs by A·U pairs. However, as shown in Figure 8, **24** can adopt three alternate conformations only one of which is similar to **1**. Free-energy calculations suggest that the conformation which is similar to **1** is less favored than the other two. For this reason, it is not surprising that **24** does not bind coat protein. It will be necessary to construct variants which cannot adopt alternate conformations to evaluate the role of the G·C pairs.

A striking feature of the coat protein binding site is the bulged A at position −10. Variant **18**, in which A_{-10} is deleted, shows little detectable interaction with coat protein, suggesting that this nucleotide is also essential for protein binding. However, since essential contacts at A_{-17} and U_{-5} are located on either side of A_{-10}, it is possible that the deletion disrupts binding by changing the distance between these two contacts. Although it was not possible to easily construct base substitutions at position −10, two variants were prepared which could potentially alter the conformation in the neighborhood of −10. As shown in Figure 8, variant **23** can potentially form an A·U base pair by using A_{-10} and have G_{-11} as a bulged nucleotide. Variant **23** binds rather well to coat protein, with a K_a only about 10-fold below that of **1**. However, since **23** can also adopt an alternative conformation with a G·U pair that resembles **1** (Figure 8), it seems more likely that the good binding observed is due to this alternate conformation and not to one with a paired A_{-10} residue.

Variant **20** contains an extra cytidine in the upper part of the helix. Either this molecule can have a conformation which places the extra nucleotide opposite A_{-10}, creating an internal loop, or it can form a five-residue loop and have the correct placement of A_{-10} (Figure 8). Since variant **20** binds poorly

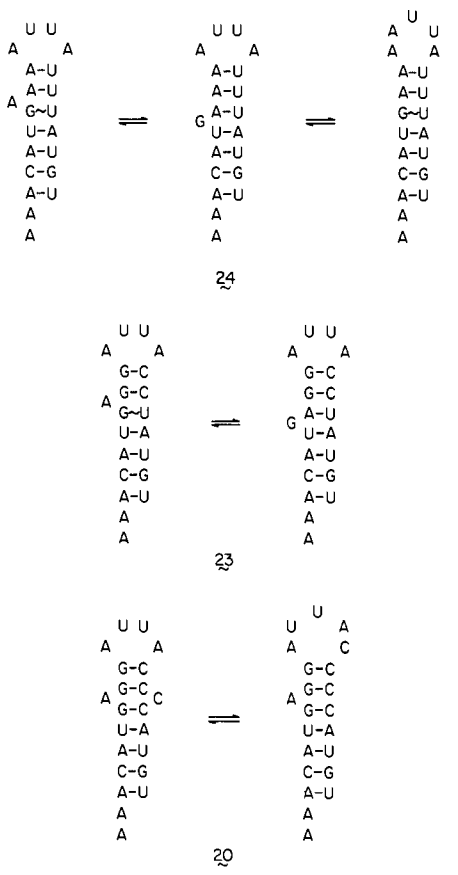

FIGURE 8: Alternate secondary structures of several variants.

to coat protein, it is likely that both these conformations disrupt the interaction with the protein. The five-residue loop in the one conformation disrupts essential contacts at −4 and −5, and the internal loop in the other conformation disrupts a contact at −10. In any case, although other variants must be made to confirm the point, it appears likely that A_{-10} must be present as a bulged nucleotide for effective protein binding.

Discussion

In the case of several DNA binding proteins, systematic analysis of protein binding to variant DNAs has been used to develop detailed models of the number and types of interactions in the protein–nucleic acid complex. A decrease in K_a for protein binding to a variant compared to the natural sequence has been interpreted as resulting from the loss or disruption of one or more stabilizing contacts between the protein and the DNA. Analysis of a variety of natural and synthetic variants of the *lac* operator enabled Caruthers (1980) to propose that the *lac* repressor protein contacts several nucleotides on one face of the operator DNA helix. Similarly, Ohlendorf et al. (1982) estimated the number of hydrogen bonds involved in contacts between *cro* protein and its primary DNA binding site in λ O_R3 from the reduced K_a's of mutant and secondary operator sites.

In the R17 case, virtually nothing was known about the coat protein binding behavior of natural variants, nor was chemical modification data available. Therefore, two general considerations governed the choice of synthetic variants made in this work. First, it was necessary to prepare a series of RNA molecules with alterations at a wide range of positions so that a general idea of the size and complexity of the binding site could be obtained. Since the binding site consists of a hairpin loop of four nucleotides with a single bulged adenine in the stem, these structural features were especially of interest. The

second major consideration in the choice of variants was the ease of synthesis. The branched scheme used in the synthesis of the native sequence was designed to make use of six easily obtained short oligomers. The protocols for variants prepared in this work avoided any major departures from the general protocol and concentrated on making changes at positions where new short oligomers could efficiently be introduced. Later experiments will attempt more complicated syntheses as they become necessary.

The binding of coat protein to each variant was determined with a nitrocellulose filter assay by varying the concentration of coat protein at a constant low concentration of labeled RNA. Although it is not an equilibrium method, this assay was previously shown to give a valid measure of the equilibrium constant for coat protein binding to the wild-type 21-mer (Carey et al., 1983). A major disadvantage of this assay is that, due to the uncertainties about coat protein aggregation and filter surface effects, the accuracy of K_a values below about 10^7 M^{-1} is unknown. Since many of the variants bind less well than 10^7 M^{-1}, it is not possible to assess the contribution of a given nucleotide to the total K_a of the interaction.

The interpretation that a decrease in the K_a of a variant is due to alteration of a contact with the protein is complicated by the structure of RNA molecules. A precise structure appears to be an important requirement for specific binding of coat protein to the natural 21-mer sequence (Carey et al., 1983). Although we know that the natural sequence adopts a hairpin loop structure as depicted in Figure 1, the secondary structure of each variant has not been determined. Since some of the variants are predicted to have altered secondary structures, a decrease in K_a may be due to structural rearrangment in these RNA molecules. We cannot distinguish this case from simple loss of a protein contact.

Although the collection of variants presented here is incomplete, the general features of the RNA binding site are relatively clear. The size of the binding site was determined to extend from A_{-17} to A_{+1} by measuring the K_a of variants shorter than the 21-nucleotide fragment. These results agree quite well with the less quantitative selection experiments of Carey et al. (1983). The binding domain is asymmetric with respect to the hairpin secondary structure of this region. It is clear that the upper portion of the hairpin loop, encompassing residues −12 to +1, is essential for protein binding. This conclusion is substantiated by the fact that variants with additions, deletions, or substitutions in this region or variants in which the upper four base pairs are disrupted all show reduced values of K_a. On the other hand, a base-paired structure in the lower part of the hairpin does not appear to be necessary for protein binding, since residues U_{+2} to U_{+4} can be deleted without affecting K_a. Since deletion of A_{-17} results in a very large decrease in K_a, it is clear that this residue is part of the binding site. Thus, taken together, the affinities of these variants can be understood in terms of protein contacts with two different locations of the RNA molecule: the upper part of the hairpin loop and the 5′-terminal A residue.

Two presumably independent sites of contact between the coat protein and a single-stranded region of the RNA were studied in greater detail by substituting each of the three common bases for the naturally occurring base. Changing the adenine at position −4 to any of the other bases reduces K_a dramatically. Although the magnitude of the decrease in K_a is not known with certainty, it seems quite large considering that only 1 base in 21 was altered. In contrast, single base pair replacements in the *lac* operator typically give relatively small (5–50-fold) reductions in K_a for the *lac* repressor (Gilbert et al., 1975; Goeddel et al., 1978). It is possible that the single-stranded structure of A_{-4} permits more extensive contact between the nucleotide and the coat protein than generally occurs with DNA binding proteins. More extensive contact could result in a much larger contribution of a single nucleotide to the total binding free energy and a greater decrease in K_a upon substitution. Another protein–nucleic acid interaction which appears to be very sensitive to single base substitution is the interaction of tRNA with its cognate synthetase. Substitution of single nucleotides in the anticodon of *E. coli* $tRNA^{Gly}$ and $tRNA_f^{Met}$ can cause greater than thousandfold reductions in the rate of aminoacylation (Squires & Carbon, 1971; Schulman & Pelka, 1977; Schulman et al., 1983).

The interaction between coat protein and the uracil at position −5 seems to be quite different than the interaction at −4. The substitution of bases for U_{-5} does not lead to as large a reduction in K_a, suggesting that fewer or less extensive contacts are made at this position. In addition, the value of K_a obtained depends upon the particular substitution made. This result is also observed at several locations in *lac* (Betz & Sadler, 1981) and *cro* (Johnson et al., 1979) operators and suggests that the substituting bases can accommodate the protein to different extents. Because each base may be the site of more than one contact with the protein, some contact sites will be only partially fulfilled by certain base substitutions. Furthermore, some base replacements could additionally destabilize the interaction by preventing contacts with other sites nearby.

It is particularly interesting that the substitution of U_{-5} by a C leads to a tighter binding variant. This result shows that the binding site sequence was selected to give a particular value of the equilibrium constant, rather than the highest possible one. An R17 mutant having a C at position −5 would make less replicase protein because translational repression would occur earlier during infection at a lower coat protein concentration.

Two other potential sites of interaction between coat protein and RNA are at A_{-10} and A_{-17}. Since these two sites have been identified only by the fact that their deletion reduces K_a considerably, much less is known about them than about sites where base substitutions were made. It is interesting to note that the analogous region on phage Qβ RNA which binds Qβ coat protein also has a bulged adenine and several 5′ single-stranded adenines at similar positions (Weber, 1976). The existence of bulged adenines in the binding sites of several RNA binding proteins has recently been noted (Peattie et al., 1981).

Four of the seven single-stranded nucleotides in the binding hairpin were therefore identified as potential sites of contact with the protein. These four residues are dispersed over the entire hairpin, implying that the protein and RNA are in contact with one another over an extensive area. We have previously calculated that the sizes of the hairpin and the coat protein monomer are approximately the same (Carey et al., 1983). Since the remaining three single-stranded nucleotides lie within the calculated contact area, they may also bind the coat protein. In addition, although we have demonstrated the importance of the hairpin secondary structure, we do not know whether contacts between the protein and the base-paired nucleotides occur. Additional variants will be required to settle these points.

The molecular details of each of the contacts identified here cannot be determined from the present results. As exemplified by the case of the *lac* operator (Goeddel et al., 1978; Caruthers, 1980) quite subtle functional group replacements are

required to infer molecular level details. The broad substrate specificity of RNA ligase (Gumport & Uhlenbeck, 1981) should permit incorporation of a wide variety of nucleotide analogues at the contact sites identified here. Such variants should also be useful for understanding the energetic consequences of altering a contact, since the more subtle substitutions are expected to show smaller changes in K_a.

Registry No. **1**, 86784-73-8; **2**, 86784-63-6; **3**, 86784-59-0; **4**, 86784-57-8; **5**, 86784-66-9; **6**, 86784-62-5; **7**, 86784-61-4; **8**, 86784-60-3; **9**, 86784-58-9; **10**, 86784-74-9; **11**, 86784-68-1; **12**, 86784-67-0; **13**, 86784-78-3; **14**, 86784-75-0; **15**, 86784-77-2; **16**, 86784-76-1; **17**, 86784-72-7; **18**, 86784-64-7; **19**, 86784-65-8; **20**, 86784-81-8; **21**, 86784-79-4; **22**, 86784-70-5; **23**, 86784-71-6; **24**, 86784-69-2; ApApCpApUpG, 86669-63-8; UpGpCpCpC, 86669-64-9; UpCpCpCpC, 86669-65-0; UpUpCpCpC, 56931-11-4; GpApCpCpC, 86669-66-1; CpApCpCpC, 86669-67-2; ApApCpCpC, 86669-68-3; ApCpCpCpC, 86669-69-4; UpApCpC, 82604-49-7; UpApCpCpCpC, 86669-74-1; UpApUpU, 86669-75-2; UpApUpUpU, 86669-76-3; UpApApApA, 86669-70-7; UpApCpCpU, 86669-71-8; UpApUpUpC, 86669-72-9; ApApApApU, 70700-49-1; pGpGpApUp, 86669-73-0; ApApCpGpGpApUp, 86688-55-3.

References

Bernardi, A., & Spahr, P. F. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3033–3037.

Betz, J., & Sadler, J. (1981) *Gene 13*, 1–12.

Carey, J., & Uhlenbeck, O. C. (1983) *Biochemistry 22*, 2610–2615.

Carey, J., Cameron, V., de Haseth, P. L., & Uhlenbeck, O. C. (1983) *Biochemistry 22*, 2601–2610.

Caruthers, M. H. (1980) *Acc. Chem. Res. 13*, 155–160.

Gilbert, W., Gralla, J., Majors, J., & Maxam, A. (1975) in *Protein–Ligand Interactions* (Sund, H., & Blauer, G., Eds.) pp 193–210, Walter de Gruyter, Berlin.

Goeddel, D. V., Yansura, D. G., Winston, C., & Caruthers, M. H. (1978) *J. Mol. Biol. 123*, 661–687.

Gumport, R. I., & Uhlenbeck, O. C. (1981) in *Gene Amplification and Analysis* (Chirikjian, J., & Papas, T., Eds.) Vol. 2, pp 313–345, Elsevier/North-Holland Biomedical Press, New York.

Jansone, I., Berzin, V., Gribanov, V., & Gren, E. J. (1979) *Nucleic Acids Res. 6*, 1747–1760.

Johnson, A., Meyer, B. J., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5061–5065.

Krug, M., deHaseth, P. L., & Uhlenbeck, O. C. (1982) *Biochemistry 21*, 4713–4720.

Lodish, H. F., & Zinder, N. D. (1966) *J. Mol. Biol. 19*, 333–348.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982) *Nature (London) 298*, 718–723.

Peattie, D. A., Douthwaite, S., Garrett, R. A., & Noller, H. F. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 7331–7335.

Schulman, L. H., & Pelka, H. (1977) *Biochemistry 16*, 4256–4265.

Schulman, L. H., Pelka, H., & Susani, M. (1983) *Nucleic Acids Res. 11*, 1439–1446.

Shafranski, P., Zagorski, V., Khrobochek, Ya., & Zagorska, L. (1975) *Mol. Biol. (Engl. Transl.) 9*, 78–85.

Squires, C., & Carbon, J. (1971) *Nature (London), New Biol. 233*, 274–277.

Steitz, J. A. (1974) *Nature (London) 248*, 223–225.

Uhlenbeck, O. C., & Cameron, V. (1977) *Nucleic Acids Res. 4*, 85–98.

Weber, H. (1976) *Biochim. Biophys. Acta 418*, 175–183.

On the Determination of Deoxyribonucleic Acid–Protein Interaction Parameters Using the Nitrocellulose Filter-Binding Assay†

Charles P. Woodbury, Jr.,* and Peter H. von Hippel

ABSTRACT: We examine the effects of filter efficiency on DNA–protein binding data obtained by the popular nitrocellulose filter-binding assay. Graphical procedures for determining the efficiency parameter ϵ (for the efficiency of retention of DNA on the filter, per bound protein molecule) are established. Filter efficiency modified formulas for determining thermodynamic binding parameters are derived for four simple prototypes of DNA–protein binding systems. The effects of experimental error on discrimination between models are considered. Finally, we discuss conceptual errors often made in calculating binding stoichiometry from filter-binding experiments and suggest a general protocol for the analysis of filter-binding data.

The nitrocellulose filter assay is a popular tool for studying protein–nucleic acid interactions. As is well-known, double-stranded DNA binds very poorly to nitrocellulose while many proteins bind quite strongly; a protein molecule can thus retain on the filter (more or less efficiently) any double-stranded DNA molecule bound to it, since "free" DNA molecules are flushed through the filter. Typically, the DNA is radioactively labeled, and the extent of protein–DNA binding is determined by liquid scintillation counting of the filter.

A basic difficulty with interpreting results of the assay is that it monitors only the amount of DNA retained on the filter rather than the actual number of protein–DNA complexes in solution. Thus dissociation of weak complexes during the course of filtration may bias the results considerably. Further, since one protein molecule suffices (in general) to bind an entire DNA molecule to the filter, the assay does not discriminate among DNA molecules carrying one, two, or more

† From the Department of Medicinal Chemistry and Pharmacognosy, University of Illinois at Chicago, Chicago, Illinois 60680 (C.P.W.), and the Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, Oregon 97403 (P.H.v.H.). *Received February 16, 1983.* Supported by National Institutes of Health Postdoctoral Fellowship GM-05669 (to C.P.W.) and National Institutes of Health Grant GM-15792 (to P.H.v.H.).

0006-2960/83/0422-4730$01.50/0