Dickerson, D., Drew, H., Connor, B., Wing, R., Frantini, A., & Kopka, M. (1982) *Science (Washington, D.C.) 216*, 475–485.
Eigner, J., & Doty, P. (1965) *J. Mol. Biol. 12*, 549–580.
Fish, S., Chen, C., Thomas, G. J., Jr., & Hanlon, S. (1983) *Biochemistry 22*, 4751–4756.
Fodor, S., Starr, P., & Spiro, T. (1985) *Biopolymers 24*, 1493–1500.
Gessner, R., Quigley, G., Wang, A., van der Marel, G., van Boom, J., & Rich, A. (1985) *Biochemistry 24*, 237–240.
Goto, S. (1984) *Biopolymers 23*, 2211–2222.
Harvey, S. (1983) *Nucleic Acids Res. 11*, 4867–4878.
Hearst, J., & Vinograd, J. (1961) *Proc. Natl. Acad. Sci. U.S.A. 47*, 1005–1014.
Kilkuskie, R. (1982) Ph.D. Dissertation, University of Illinois, Chicago, IL.
Kilkuskie, R., Wood, N., & Hanlon, S. (1982) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 41*, 1023.
Maibenco, D., Chang, S., Chen, C., & Hanlon, S. (1984) *Biochemistry 23*, 3373, Abstract 145.
Miller, F., Dixon, G., Rattner, J., & van der Sande, J. (1985) *Biochemistry 24*, 102–109.
Mirav, P., & Kearns, D. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1594–1598.
Nickol, J., Behe, M., & Felsenfeld, G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1771–1775.
Nordheim, A., Hao, W., Wogan, G., & Rich, A. (1983) *Science (Washington, D.C.) 219*, 1434–1436.
Olson, W., Srinivasan, A., Marky, N., & Balaji, N. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 229–241.
Pack, G., & Klein, B. (1984) *Biopolymers 23*, 2801–2823.
Pack, G., Prasad, C., Salatsky, J., & Wong, L. (1986) *Biopolymers 25*, 1697–1715.
Pohl, F. (1976) *Nature (London) 260*, 365–366.
Pohl, F., & Jovin, T. (1972) *J. Mol. Biol. 67*, 375–396.
Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., & Klug, A. (1984) *Nature (London) 311*, 532–537.
Ringquist, S. (1987) Ph.D. Dissertation, University of Illinois, Chicago, IL.
Semin, Y., Volkov, V., Protasova, I., & Poverennyi, A. (1981) *Mol. Biol. 15*, 1303–1314.
Simpson, R., & Kunzler, P. (1979) *Nucleic Acids Res. 6*, 1387–1415.
Simpson, R., & Shindo, H. (1980) *Nucleic Acids* Res. *8*, 2093–2103.
Svedberg, T., & Pederson, K. D. (1940) *The Ultracentrifuge*, Clarendon, London.
Thamann, T., Lord, R., Wang, A., & Rich, A. (1981) *Nucleic Acids Res. 9*, 5443–5457.
Thomas, T. J., & Bloomfield, V. A. (1983) *Nucleic Acids Res. 11*, 1919–1930.
Thomas, T. J., & Bloomfield, V. A. (1985) *Biochemistry 24*, 713–719.
Tunis, M. J., & Hearst, J. E. (1968) *Biopolymers 6*, 1345–1353.
Vorlickova, M., Minyat, E., & Kypr, J. (1984) *Biopolymers 23*, 1–4.
Wang, A., Quigley, G., Kolpak, F., Crawford, J., van Boom, J., van der Marel, G., & Rich, A. (1979) *Nature (London) 282*, 680–686.

Role of a Bulged A Residue in a Specific RNA–Protein Interaction†

Huey-Nan Wu and Olke C. Uhlenbeck*
Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309
Received May 21, 1987; Revised Manuscript Received August 4, 1987

ABSTRACT: The translational operator of the R17 replicase gene contains a bulged A residue that is essential for the specific binding to R17 coat protein. A large number of operator variants have been synthesized to more precisely examine the role of the bulged A residue on this specific protein–RNA interaction. By use of RNA ligase and transcription of synthetic DNA templates by T7 RNA polymerase, 14 different nucleotides were introduced to the bulged A position of three different coat protein binding fragments. The affinity between coat protein and each fragment was determined by a nitrocellulose filter binding assay. The data indicate that while functional groups on N^1, C^2, C^6, N^7, and 2′OH of the bulged A can be substituted without greatly changing protein binding, bulky substituents cannot be tolerated at these positions. Data from additional fragments that have base-pair changes adjacent to the bulged A suggest that the propensity of the bulged A to intercalate into the helix can affect protein binding.

The coat protein of bacteriophage R17 binds specifically to a hairpin loop in the initiation region of the replicase gene and thereby represses translation of that gene late in phage infection (Bernardi & Spahr, 1972). This highly specific RNA–protein interaction has been extensively characterized by using a synthetic 21-nucleotide fragment corresponding to the −17 to +4 position of the replicase gene that has the same affinity for coat protein as the 3.5-kilobase R17 genomic RNA (Krug et al., 1982; Carey & Uhlenbeck, 1983). By measurement of the affinity of coat protein with more than 50 enzymatically synthesized variants of the 21-nucleotide fragment, the RNA structural requirements for tight protein binding were determined (Carey et al., 1983b; Romaniuk et al., 1987). Among the essential residues for the coat protein binding is an extrahelical or "bulged" A residue at position −10 of the replicase gene. Either deletion of this residue or substitution with a C residue reduced the K_a to coat protein by at least 1000-fold.

Peattie et al. (1981) have proposed a special role for the

†Supported by the National Institutes of Health (GM36944).
*Author to whom correspondence should be addressed.

0006-2960/87/0426-8221$01.50/0

bulged A residue in protein binding sites on RNA molecules. Bulged A residues appear in several other RNA phage translational operators with very different sequences (Weber, 1975; Inokuchi et al., 1986) as well as putative ribosomal protein binding sites on *Escherichia coli* 16S and 23S rRNA (Noller, 1984). In addition, carboxylation (Peattie et al., 1981) or deletion (Christiansen et al., 1985) of the phylogenetically conserved bulged A residue at position 66 of *E. coli* 5S RNA reduces its affinity to the ribosomal protein L18. The frequent appearance of bulged A residues in protein binding sites has led to the suggestion that, in analogy to many dehydrogenases (Rossmann, 1975), a specialized domain is present in the binding proteins that can interact specifically with functional groups on the bulged A nucleotide.

In this work the bulged A residues of three RNA fragments that bind R17 coat protein tightly are studied in detail. In each case the A residue was substituted with 13 different nucleotides, and the coat protein binding properties were determined in order to find out whether any of the functional groups of the bulged A residue can be implicated in protein binding. In addition, the sequence of the base pairs adjacent to the bulged A have been systematically changed in order to study the possible influence of base stacking on the bulged A residue.

Materials and Methods

Enzymes. RNA ligase (Moseman-McCoy et al., 1979) and polynucleotide kinase (Cameron & Uhlenbeck, 1978) were purified from T4-infected *E. coli*. Polynucleotide kinase lacking the 3′ phosphatase activity was purified from *pse*T1 T4-infected *E. coli* (Soltis & Uhlenbeck, 1982). Primer-dependent polynucleotide phosphorylase was purified from *Micrococcus luteus* (Klee, 1971). Bacterial alkaline phosphatase was purchased from Sigma. Ribonucleases A (90 Kunitz units/mg) and T1 (5000 units/mg) were purchased from Calbiochem-Behring. Ribonucleases T2 (1070 units/mg) and P1 (160 units/mg) were purchased from Sigma. T7 RNA polymerase was purified from a recombinant *E. coli* strain according to the method of Davanloo et al. (1984).

RNAs. Ribonucleoside di- and triphosphates, GpC, ApG, GpU, 7-deazaadenosine (deaza^{7}A), and 2′-*O*-methyladenosine (2′OmA) were purchased from Sigma. pAp, pGp, pCp, pUp, 1-methyladenosine (m^{1}A), and N^6-(6-aminohexyl)adenosine (AHA) 3′,5′-bisphosphate were purchased from P-L Biochemicals. ϵ-Adenosine (ϵA) 3′,5′-bisphosphate was a gift of N. J. Leonard. The 3′(2′),5′-bisphosphates of inosine (I), 6-methyladenosine (m^{6}A), 6,6-dimethyladenosine (m$_2^6$A), purine ribonucleoside (PuR), and 2-aminopurine ribonucleoside (n^{2}Pu) were gifts of W. Wittenberg. 7-Deaza-, 2′-*O*-methyl-, and 1-methyladenosine 3′(2′),5-bisphosphates were synthesized from individual nucleoside with pyrophosphoryl chloride (Sigma) by the method of Barrio et al. (1978). [γ-^{32}P]ATP was prepared from [^{32}P]orthophosphate (Johnson & Walseth, 1979). [5′-^{32}P]pCp was prepared from [γ-^{32}P]ATP (England et al., 1980). [α-^{32}P]CTP (1 Ci/mmol) was purchased from Du Pont.

Oligoribonucleotide Synthesis. Two buffers were used in most of the reactions described below. Buffer A was used for RNA ligase and polynucleotide kinase reactions and contains 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 8.3, 20 mM $MgCl_2$, and 3 mM dithiothreitol. Buffer B was used for T7 RNA polymerase transcription reactions and contains 40 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), pH 8.0, 6 mM $MgCl_2$, 1 mM spermidine, 5 mM dithiothreitol, and 0.1 mM bovine serum albumin (BSA).

Oligonucleotides less than 13 residues were purified from reaction mixtures by ion-pair HPLC (Beckett & Uhlenbeck, 1984) and desalted by HPLC as described by Weber (1985). Longer oligonucleotides were purified by gel electrophoresis, eluted from gel slices, and ethanol precipitated (Krug, 1982).

Fourteen different octamers, A_6CN, present at the 5′ end of fragments **1-N**, **2-N**, and **3-N** were each synthesized by the same procedures. A_6C was prepared by RNase A digestion of AC copolymer followed by treatment with alkaline phosphatase (Uhlenbeck & Cameron, 1977). To synthesize each A_6CNp, a 200-μL reaction containing 0.1 mM A_6C, 1 mM nucleoside 3′(2′)-bisphosphate (pNp), 1.5 mM ATP, and 300 units/mL RNA ligase in buffer A was incubated at 14 °C for 20 h. The reaction was terminated by heating the reaction at 90 °C for 3 min, and alkaline phosphatase was added to 0.1 mg/mL and incubated at 37 °C for an additional 20 h. After purification, about 10 nmol of each A_6CN was recovered.

GpCpApApCpApGpCpGpUpUpUpU, the common 3′ tridecamer of fragments **1-N**, was synthesized in a stepwise fashion analogous to the procedures described previously by Krug et al. (1982), Carey et al. (1983b), and Beckett and Uhlenbeck (1984). GpCpApA and GpUpUpUpU were prepared from equilibrium polynucleotide phosphorylase reactions. ApGpC was prepared from RNase A assisted polynucleotide phosphorylase reaction followed by treatment with alkaline phosphatase. GpCpApApC was synthesized from GpCpApA and pCp in a similar manner to the synthesis of A_6CN described above. pGpUpUpUpU was made by incubating 0.5 mM pentamer, 1.0 mM ATP, and 100 units/mL polynucleotide kinase in buffer A at 37 °C for 2 h. The yield of purified oligomer was about 90%. ApGpCpGpUpUpUpU was prepared in a 1.2-mL reaction containing 0.53 mM AGC, 0.33 mM pGpUpUpUpU, 0.5 mM ATP, and 200 units/mL RNA ligase in buffer A and incubated 20 h at 14 °C. A total of 250 nmol (60%) of HPLC-purified octamer was obtained. The octamer was 5′ phosphorylated in a reaction the same as for the synthesis of pGpUpUpUpU, and the reaction was terminated by heating to 90 °C for 3 min. The kinase reaction mixture was added directly to an RNA ligase reaction containing 60 μM pAGCGUUUU, 150 μM GpCpApApC, 150 mM ATP, and 200 units/mL RNA ligase in buffer A. After incubation at 14 °C for 20 h, the tridecamer was purified and desalted by HPLC, and the recovered yield was 35%.

GpCpApUpCpApGpCpGpUpUpUpUp and GpGpApUpCpApCpCpGpUpUpUpUp, the 3′ tridecamers of fragments **2-N** and **3-N**, respectively, were synthesized with the aid of in vitro transcription of synthetic DNA templates by T7 RNA polymerase (Milligan et al., 1987). Template strand DNA fragments 3′-ATTATGCTGAGTGATATCGTAGTCGCAAA and 3′-ATTATGCTGAGTGATATCCTAGTGGCAAA were synthesized on an Applied Biosystems DNA synthesizer and purified by polyacrylamide gel electrophoresis (Caruthers, 1985). Each was annealed to the 18-mer 5′-TAATACGACTCACTATAG, which corresponds to −17 to +1 of the T7 promoter, to give the partially double-stranded templates required by T7 polymerase (Milligan et al., 1987). Annealing was carried in 10 mM Tris-HCl, pH 7.8, 5 mM $MgCl_2$, and 0.1 mM ethylenediaminetetraacetic acid (EDTA) at 25 °C for 30 min. The tridecamers were prepared in three steps. First, the dodecamers pppGpCpApUpCpApGpCpGpUpUpU and pppGpGpApUpCpApCpCpGpUpUpU were synthesized in transcription reactions consisting of 500 nM DNA templates, 60 units/μL T7 RNA polymerase, and 700 μM each of four rNTP in buffer B in 10 mL. After incubation at 37 °C for 1 h, the reaction

mixtures were extracted successively with 10 mL of phenol/chloroform (1:1 v/v) and 10 mL of chloroform/isoamyl alcohol (24:1 v/v). The transcripts were then ethanol precipitated and purified by gel electrophoresis. The relatively short length of the T7 transcription templates necessitated higher template and enzyme concentrations than used normally (Milligan et al., 1987) and resulted in modest yields (20 nmol of product). The second step involved the removal of the 5′ terminal triphosphate by incubating each transcript in 0.1 mg/mL alkaline phosphatase, 10 mM Tris-HCl, pH 8.0, and 0.1 mM EDTA and repurifying the dephosphorylated products by gel electrophoresis. In the third step the dodecamers were reacted with pUp and RNA ligase by the same protocol described in the synthesis of A_6CNp.

Internally ^{32}P-labeled fragments **1-N**, **2-N**, and **3-N** were prepared in two steps. A 10-μL reaction medium containing 20 μM each 3′ $^3/_4$ molecule, 30 μM [γ-^{32}P]ATP (160 Ci/mmol), and 100 units/mL either wild type (for **1-N**) or *pse*T1 (for **2-N** and **3-N**) polynucleotide kinase in buffer A. After incubating at 37 °C for 2 h, the reaction was terminated by heating at 90 °C for 3 min. Then a 20-μL reaction containing 10 μL of the kinase reaction, 20 μM A_6CN, 1 mM ATP, and 250 units/mL RNA ligase in buffer A was incubated at 14 °C for 14 h. Fragments **1-N**, **2-N**, and **3-N** were gel purified, eluted from gel slices, and ethanol precipitated.

The absence of a 3′ terminal phosphate on GpCpApApCpApGpCpGpUpUpUpU results in several undesirable side reactions, including intramolecular cyclization and dimerization. Although the side products could be separated from the correct product in each reaction, the overall yield of fragments **1-N** was considerably less than that of **2-N** and **3-N** where the donors contained 3′ terminal phosphates.

The ^{32}P-labeled fragments **4–16** were synthesized by in vitro transcription reactions with T7 RNA polymerase using the appropriate 41-nucleotide template strand annealed to the 18-nucleotide T7 promoter strand described above. Each 20-μL transcription reaction in buffer B contained 50 nM template, 100 μM [α-^{32}P]CTP (12.5 Ci/mmol), 700 μM each of UTP, ATP, and GTP, and 60 units/μL T7 RNA polymerase. After incubation at 37 °C for 1 h, transcripts (fragments **4–16**) were purified by gel electrophoresis. Lowering the CTP concentration with respect to the other triphosphates increased the yield of radiolabeled product (1.5 μCi).

Product Identification. Procedures for the identification of GpCpApApCpApGpCpGpUpUpUpU and its intermediates were analogous to those used for a variety of other RNA fragments made by branched RNA ligase pathways (Krug et al., 1982; Carey et al., 1983b). The sequences of the two dodecamers made by T7 transcription were confirmed by preparing the molecules with either [α-^{32}P]UTP or [α-^{32}P]CTP and showing that the expected digestion products were obtained with RNase A, RNase T1, or RNase T2 (Milligan et al., 1987). Each internally ^{32}P-labeled fragment **1-N**, **2-N**, and **3-N** was shown to contain the correct internucleotide linkage at the final joining point by digestion with either nuclease P1 to give the labeled nucleoside 5′-monophosphate present on the 5′ terminus of the tridecamer or mixed nucleases to give the labeled nucleoside 3′-monophosphate present on the 3′ terminus of the octamer (Carey et al., 1983b).

Coat Protein Binding. R17 coat protein was prepared by the method of Beckett and Uhlenbeck (1988). The association constant between the coat protein and each fragment was determined by a nitrocellulose filter retention assay (Carey et al., 1983a). For each fragment, approximately 3 nCi of oligomer was mixed with 12 different coat protein concentrations ranging from 0.1 nM to 1 μM in 10 mM magnesium acetate, 80 mM KCl, 80 μg/mL serum albumin, and 100 mM Tris-HCl, pH 8.5. After incubation at 2 °C for 1.5 h, the mixture was filtered through a nitrocellulose filter and the amount of complex retained on the filter determined by liquid scintillation counting. For each experiment the data points were fit to a filter retention efficiency and then to a K_a value assuming a bimolecular equilibrium. The coat protein concentration is expressed in moles of coat protein dimer since it is the interacting species (Beckett & Uhlenbeck, 1988). Duplicate assays using the same set of coat protein dilutions give K_a values within a factor 2. The precision in K_a for independent coat protein dilutions is a factor 3. Each reported K_a value is the average of 2 to 10 independent determinations.

Association and dissociation rates between coat protein and fragments **3-A**, **3-deaza^{7}A**, **3-2′OmA**, and **3-PuR** were determined as described in Carey and Uhlenbeck (1983).

Results

Synthesis of RNA Fragments. The 4 regular and 10 modified nucleotides were introduced into the bulged A position of the 21-nucleotide R17 coat protein binding fragment by an enzymatic procedure using a heptamer, $(Ap)_6C$, a tridecamer, and a series of nucleoside 3′(2′),5′-bisphosphates. The addition of each pNp to $(Ap)_6C$ by RNA ligase was accomplished in very high yields. After removal of the 3′ terminal phosphate and purification by HPLC, the resulting octamers were used as acceptors in a second RNA ligase reaction with a 5′ ^{32}P-labeled tridecamer donor. The resulting internally labeled 21-mers were purified by gel electrophoresis and used directly in protein binding assays. As a consequence of the position of the ^{32}P label, the presence of the appropriate modified nucleotide could be confirmed by nuclease digestion and thin-layer chromatography.

Three different tridecamers were used as donors in the second RNA ligase reaction, resulting in different "background" sequences for the bulged A variants. The tridecamers used for fragments **1-N** were prepared by the well-described stepwise joining of shorter oligomers with RNA ligase. The multiple steps resulted in only small amounts of product being obtained. The tridecamers used for fragments **2-N** and **3-N** were made by transcription of synthetic DNA by T7 RNA polymerase. Although the sequence and lengths of the templates were not favorable for higher transcription yields (Milligan et al., 1987), much more material was obtained by this method than the former one, and the effort was far less. As a result of the availability of more material, the 3′ termini of the fragments used in **2-N** and **3-N** were blocked by the addition of a pUp. This resulted in much higher yields of 21-mer in the final ligation step.

Fragments **4–16** (Figure 1) were all prepared by transcription of synthetic DNA templates. Since these fragments were longer and the first five nucleotides corresponded precisely to the type III T7 promoter, very high yields of products were obtained.

Substitutions of the Bulged A Residue. The sequence of the R17 replicase translational operator (**O**) is substantially different from **1-A**, **2-A**, **3-A**, and **4**, the fragments used as controls in this work (Figure 1). Nevertheless, as seen in Table I, the K_a values of these four fragments to R17 coat protein compare sequences satisfy the combination of sequence and structure proposed to be necessary for coat protein binding (Romaniuk et al., 1987).

The coat protein binding affinities of fragments **1**, **2**, and **3** with a U, C, or G in place of the A are listed in Table I. In all cases where a pyrimidine was substituted, the K_a is

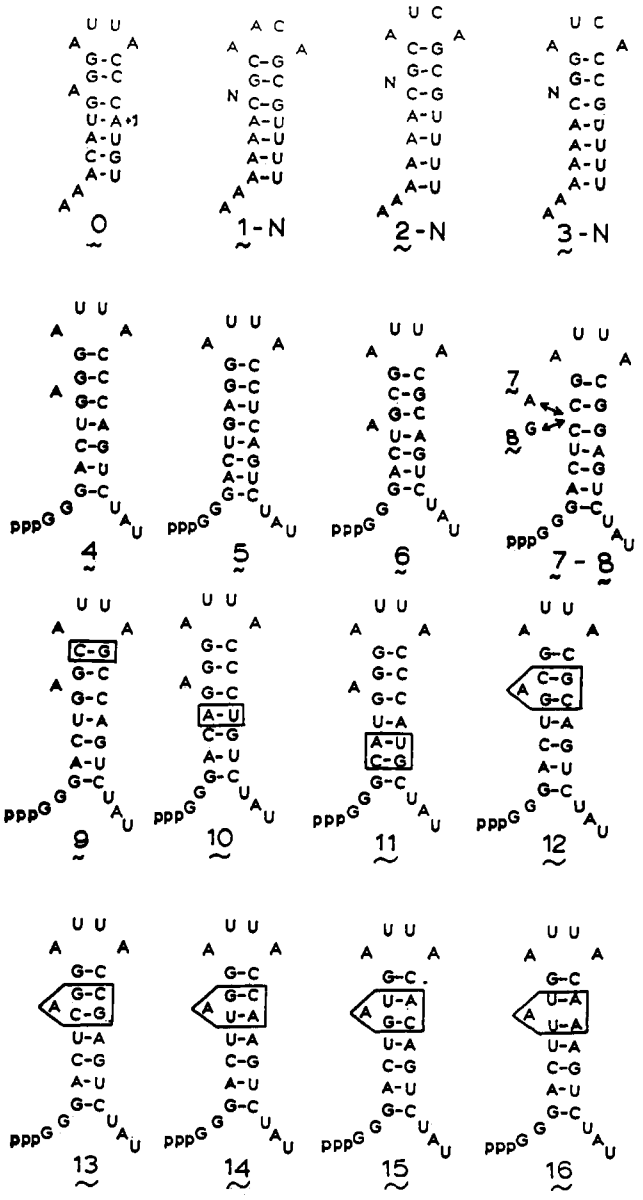

FIGURE 1: Sequences and possible structures of fragments. N indicates positions where different nucleotides have been inserted. The nucleotides in fragments **9**–**16** that are different from those of **4** are boxed.

Table I: Coat Protein Binding to RNA Fragments with Regular Nucleotides at the Bulged Position

fragment	$10^8 K_a$ (M^{-1})	fragment	$10^8 K_a$ (M^{-1})
0	6.0	**3-A**	304[a]
1-A	6.4	**3-G**	<0.02
1-G	1.3	**3-U**	0.29
1-U	<0.02	**3-C**	<0.02
1-C	<0.02	**4**	10.4
2-A	19[a]	**5**	<0.002
2-G	3.8	**6**	<0.002
2-U	<0.02	**7**	2.1
2-C	<0.02	**8**	3.6

[a] Data calculated from k_{on}/k_{off}.

reduced at least 1000-fold. This is in agreement with the observation that coat protein binding was eliminated when the bulged A was changed to C in the wild-type sequence (Romaniuk et al., 1987). However, the results for substitution of a G for the A depend upon the background sequence. Fragments **1-G** and **2-G** have only a 5-fold lower K_a than their A-containing counterparts, whereas **3-G** binds more than 1000-fold less well than **3-A**.

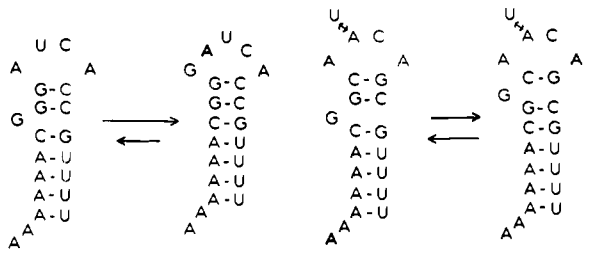

FIGURE 2: Structure of variants with a G at the bulged position.

Table II: Coat Protein Binding to RNA Fragments with Modified Nucleotide Substitutions of the Bulged A Residue

N	relative K_a for background sequence 1	2	3
A	1[a]	1[b]	1[c,d]
G	0.2	0.2	<0.0001
n^2Pu	1.1	0.29	<0.008
I	1.5	0.42	0.001
deaza^{7}A	0.54	0.65	0.8[d]
2′OmA	0.57	0.31	0.8[d]
m^6A	0.1	0.04	0.009
m^2_6A	<0.003	0.0052	0.0009
m^1A	<0.0003	<0.0001	0.002
ϵA	<0.003	<0.001	0.0005
AHA	0.01	0.025	0.003
PuR	0.19	0.12	0.11[d]

[a] $K_a = 6.4 \times 10^8$ M^{-1}. [b] $K_a = 1.9 \times 10^9$ M^{-1}. [c] $K_a = 3.0 \times 10^{10}$ M^{-1}. [d] K_a calculated from k_{on}/k_{off}.

A likely explanation for the different effect of the A to G substitution when different background sequences are used is that the G-substituted fragments can take on alternate conformational forms. In particular, as shown in Figure 2, fragment **3-G** would be likely to be more stable as an uninterrupted helix and a five base pair loop, whereas fragments **1-G** and **2-G** could also form structures with the bulged nucleotide in a different position. Since these alternate structures do not have the structural requirement for coat protein binding, their propensity to form could reduce the K_a. Since the increased stability of the alternate structure is likely to be greatest for **3-G**, it is not surprising that it binds coat protein the least well.

In order to test this explanation and to detect the real effect of the A to G substitution on coat protein binding, fragments **7** and **8** (Figure 1) were constructed. Both fragments have the same background sequence and are unlikely to form alternative structures. As shown in Table I, fragments **7** and **8** bind coat protein equally well. This suggests that either purine is equally suitable for coat protein binding at this position as long as the surrounding sequence permits the correct structure. It should be noted that although fragments **1-C**, **2-C**, and **3-C** could form similar alternate structures, it is unlikely that a C can be accommodated at the bulged A position because the fragment constructed by Romaniuk et al. (1987) with a C at position −10 could not form alternate structures and has an undetectable affinity to coat protein. Taken together, the data suggest that the essential contacts for coat protein binding at position −10 can be provided by either purine, but pyrimidines cannot be accommodated.

Two additional variants were constructed to further explore the role of the bulged purine. Fragment **6** has the position of the bulged A residue moved down one base pair in the helix. Fragment **5** introduces a U opposite the bulged A, permitting a continuous helix to form. Both of these variants show nearly no detectable binding to coat protein. Thus, the position and conformation of the bulged A residue is also important for proper interaction with the protein.

Variants with Modified Bases in the Bulged Position. The values of K_a between coat protein and variants **1**, **2**, and **3** with 10 different modified nucleotides at the bulged position are given in Table II normalized to fragments **1-A**, **2-A**, and **3-A**. If the contribution of the modified nucleotide to the total free energy of binding is the same for each background sequence, one would expect that the normalized K_a's would also be the same. Indeed, for 8 of the 10 modified nucleotides, the normalized K_a's for **1-N**, **2-N**, and **3-N** are within a factor of 10 of each other. Considering an error of a factor 3 in the determination of K_a and a nearly 50-fold range of the K_a's of the three background sequences, this is quite good agreement. The only data that are not consistent among the three background sequences are the K_a values of **3-n²Pu** and **3-I**, which are 100–1000 times lower than those of the same modifications in **1** and **2**. It is likely that these differences can be attributed to the alternate pairing arrangement for **3** that was discussed above. Both n^2Pu and I can pair with C much better than the other eight modified purines tested. We can therefore conclude that the effect of each nucleotide substitution on protein binding is the same for all three background sequences, thus increasing our confidence in the data.

As shown in Table II, five derivatives of the bulged A residue (m^6A, m_2^6A, m^1A, AHA, and ϵA) all bind coat protein very poorly. Since all of these have a bulky substituent at the amino and/or N1 positions of A, the data suggest a mode in which these two positions fit into a pocket in the coat protein and form specific hydrogen bonds with amino acid side chains.

However, the remainder of the data in Table II indicates that such a simple model is unlikely to be correct. Since, as discussed above, G can substitute for A without altering K_a, a simple hydrogen-bonding model will not suffice since the functional groups at N^1, C^2, and C^6 are all different between the two purines. This is supported by the fact that the n^2Pu and I derivatives, which only have some of the differences between A and G, also bind protein tightly. It is, however, surprising that the Pu derivative that is missing functional groups at C^2 and C^6 (but is N^1–H) binds coat protein significantly less well than A. The tight binding of the deaza^{7}A derivative eliminates the attractive possibility that the N^7 was used as a hydrogen-bond acceptor to distinguish purines from pyrimidines at the bulged A position. Binding of the 2′OmA derivative suggests that the ribose ring is also not involved in specific contact with the protein.

Variants with Base-Pair Changes. NMR experiments on deoxyoligoribonucleotides with a bulged A residue clearly indicated that the adenosine ring is stacked between the adjacent base pairs, much as a dye molecule intercalates into a helix (Patel et al., 1982). While it is not certain that the same situation occurs when the bulged A in the RNA hairpin is bound to coat protein, it is possible that stacking interactions between the bulged A and adjacent base pairs may influence binding to coat protein. Although previous studies have shown that changing the identity of base pairs in the hairpin had very little effect on protein binding (Romaniuk et al., 1987), it seemed worthwhile to investigate this possibility in greater detail. Six variants of the coat protein binding site were prepared that changed the base pairs around the bulged A without introducing possible alternate conformational forms (Figure 1). Three control variants were made where base pairs not adjacent to the bulged A were changed. The affinity of these variants to R17 coat protein are given in Table III. Variants **7** and **12–16**, which have base pairs changed adjacent to the bulged A, all have K_a values slightly less than **4**, the control sequence. On the other hand, variants **9–11**, which change one or more of the other base pairs, have K_a's the same as that for variant **4**. These data suggest that neighbors to the bulged A have a small effect on K_a, presumably due to stacking interactions influencing the ability of the bulged A to intercalate.

Discussion

The enzymatic syntheses of fragments **2-N** and **3-N** are examples of a general three-step protocol for the preparation of RNA molecules that contain a single modified nucleotide at any predefined site on the molecule. The first step involves the synthesis of the two RNA fragments containing the four normal nucleotides that lie on either side of the desired modification. In the case of **2-N** and **3-N**, the 5′ fragment was made with polynucleotide phosphorylase and the 3′ fragment with T7 RNA polymerase. However, in general, both these RNA fragments would be prepared by in vitro transcription with T7 RNA polymerase since large amounts of virtually any sequence can be made by this method (Milligan et al., 1987). Although the optimal reaction conditions can vary depending upon the sequence and structure of the acceptor RNA fragment, this step can generally be carried out in very high yield. The second step involves the addition of the modified nucleotide in the form of a 3′,5′-bisphosphate to the 3′ end of the 5′ fragment using T4 RNA ligase followed by treatment with alkaline phosphatase to remove the 3′ phosphate. A wide variety of 3′,5′-bisphosphates have been shown to be active in this reaction (Barrio et al., 1978). The third step involves the addition of the 5′ phosphorylated 3′ fragment to the modified 5′ fragment using RNA ligase a second time. As was found in this work, it is prudent to block the 3′ terminus of the 3′ donor fragment to prevent its cyclization with RNA ligase. In most cases this last step also can be carried out in very high yields, although ligation points in the middle of highly structural regions are to be avoided.

In most respects this modification protocol follows that used in this laboratory for altering uridine-33 in yeast $tRNA^{Phe}$ (Wittenberg & Uhlenbeck, 1985). The major difference is that in that work the 3′ and 5′ terminal fragments were generated by partial nuclease digestion experiments, and here they were individually synthesized. This represents an important difference since not all sites on natural tRNAs are available for a unique cleavage step. Thus, a considerably greater flexibility is available when an entirely synthetic approach is used.

Several experiments in this paper confirm earlier experiments showing that the bulged A residue is an essential residue for coat protein binding. However, a G can also be accommodated at this position provided certain surrounding nucleotides are also changed so that alternate conformations cannot form. This somewhat surprising result emphasizes two important points about performing structure–function experiments with RNA. First, a nucleotide should be substituted by each of the other three nucleotides before concluding whether or not the position is essential. At the four "specific" positions in R17 coat protein binding fragment several different responses to nucleotide substitution are observed. At two positions (−4 and −7) only A can be accommodated. At position −5, either pyrimidine is acceptable, and at position −10, either purine is acceptable. Thus, unless all three nucleotides are substituted, it is possible to either miss essential nucleotides or incompletely determine the degree of specificity. The second important point is that single nucleotide substitutions can cause the RNA molecule to fold in a significantly different way and thus negate the relevant conclusion of the substitution. This was seen when G was substituted for A in

Table III: Coat Protein Binding to RNA Fragments with Base-Pair Changes

fragment	10^8K_a (M^{-1})	fragment	10^8K_a (M^{-1})
4	10.4	**7**	2.1
9	8.0	**12**	3.0
10	9.2	**13**	2.5
11	8.8	**14**	2.7
		15	1.1
		16	0.8

the **3** background sequence. It was necessary to make four additional nucleotide substitutions in order to clearly demonstrate that G could fit in the bulged A position. This emphasizes the limitations of a genetic approach to structure–function studies where single nucleotide substitutes are generally obtained.

Interpretation of coat protein binding to variants involving the bulged A residue is greatly complicated by the propensity of the bulged nucleotide to intercalate into the helix. Patel et al. (1982) and Hare et al. (1986) have shown that an extra helical A between GC pairs in DNA will intercalate and cause substantial local perturbation in the helical structure. In contrast, an extrahelical C between AT base pairs is bulged out and disrupts the helix much less (Morden et al., 1983). Thus, Morden et al. (1983) suggest that both the identity of the extrahelical base and the sequence of base pairs on each side determine whether the base remains in the helix or is bulged out. Since it seems likely that a similar conclusion can be made for bulged nucleotides in RNA, this can explain the observation that the K_a of protein binding depends on the sequence of base pairs adjacent to the bulged A. For the variants that bind less well, some of the free energy of protein binding is required to shift the A residue into a configuration needed for binding. It is interesting to note that the variant that binds the best has two strongly stacking GC base pairs on either side of the bulged A, while the variant that binds the least well has adjacent weakly stacking UA pairs. It is therefore tempting to speculate that the bulged A is in an intercalated configuration when it binds coat protein.

Since each modified nucleotide will have a differing propensity to intercalate, the interpretation of the data involving variants **1-N**, **2-N**, and **3-N** could be complicated by the intercalation equilibrium. However, considering the magnitude of the effects on K_a, it is possible to eliminate the model proposed in Romaniuk et al. (1987) that takes the extreme view that no contacts occur between the functional groups of the bulged A and coat protein at all and the value of K_a for different −10 variants is entirely governed by the ability of the nucleotide to intercalate. It seems highly unlikely that adding one or two methyl groups to the adenosine ring will change the intercalation equilibrium sufficiently to reduce K_a 1000-fold. A model that is consistent with most of the data assumes that although the bulged A is in an intercalated configuration, part of the nucleotide interacts directly with the protein. The introduction of bulky functional groups will block these contacts and greatly reduce the K_a. However, substitution of the A residue with I, n^2Pu, or even G is sterically less disruptive, and the potential loss of one or more hydrogen bond contacts may only have a small effect on K_a that was not detected. It is also possible that a portion of the bulged A residue fits into a hydrophobic pocket in the coat protein and no specific hydrogen bonds form.

This work emphasizes the difficulties encountered when interpreting detailed functional data in the absence of correspondingly detailed structural information. Recent advances in RNA synthesis technology (Milligan et al., 1987) make it possible to obtain enough material to analyze RNA structure by higher resolution techniques. In addition, the availability of a highly ordered crystal of the closely related MS2 virus (Valagård et al., 1986) promises to make future analysis of this system very worthwhile.

REFERENCES

Barrio, J. R., Barrio, M.-C. G., Leonard, N. J., England, T. E., & Uhlenbeck, O. C. (1978) *Biochemistry 17*, 2077–2081.

Beckett, D., & Uhlenbeck, O. C. (1984) in *Oligonucleotide Synthesis: A Practical Approach* (Gait, M. J., Ed.) Chapter 8, pp 185–197, IRL, Oxford, U.K.

Beckett, D., & Uhlenbeck, O. C. (1988) *J. Mol. Biol.* (submitted for publication).

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

Cameron, V., Soltis, D., & Uhlenbeck, O. C. (1978) *Nucleic Acids Res. 5*, 825–833.

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

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

Carey, J., Lowary, P. T., & Uhlenbeck, O. C. (1983b) *Biochemistry 22*, 4723–4730.

Caruthers, M. H. (1985) *Science (Washington, D.C.) 230*, 281–285.

Christiansen, J., Douthwaite, S. R., Christensen, A., & Garrett, R. A. (1985) *EMBO J. 4*, 1019–1024.

Davanloo, P., Rosenberg, A., Dunn, J., & Studier, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2035–2039.

England, T. E., Bruce, A. G., & Uhlenbeck, O. C. (1980) *Methods Enzymol. 65*, 65–74.

Hare, D., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7456–7464.

Inokuchi, Y., Takahashi, R., Hirose, T., Inayama, S., Jacobson, A. B., & Hirashima, A. (1986) *J. Biochem. (Tokyo) 99*, 1169–1180.

Johnson, R. A., & Walseth, T. F. (1979) *Adv. Cyclic Nucleotide Res. 10*, 135–167.

Klee, C. B. (1971) *Proced. Nucleic Acid Res. 2*, 896–911.

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

Milligan, J., Groebe, D., Witherell, G., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* (in press).

Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I., Jr. (1983) *Biochemistry 22*, 5557–5563.

Mosenman-McCoy, M. I., Lubben, T. H., & Gumport, R. I. (1979) *Biochim. Biophys. Acta 562*, 149–161.

Noller, H. (1984) *Annu. Rev. Biochem. 53*, 119–162.

Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslauer, K. (1982) *Biochemistry 21*, 445–451.

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

Romaniuk, P., Lowary, P., Wu, H. N., Stormo, G., & Uhlenbeck, O. C. (1987) *Biochemistry 26*, 1563–1568.

Rossmann, M. G., Lilgas, A., Branden, C.-I., & Banazak, L. J. (1975) *Enzymes (3rd Ed.) 9*, 62–103.

Soltis, D., & Uhlenbeck, O. C. (1982) *J. Biol. Chem. 257*, 11332–11339.

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

Valegård, K., Unge, T., Motelins, I., Strandberg, B., & Fiers, W. (1986) *J. Mol. Biol. 190*, 587–591.
Weber, K., & Konisberg, W. (1975) in *RNA Phages* (Zinder, N. D., Ed.) pp 51–84, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Weber, T. (1985) Ph.D. Thesis, University of Illinois.
Wittenberg, W. T., & Uhlenbeck, O. C. (1985) *Biochemistry 24*, 2705–2712.

Site and Sequence Specificity of the Daunomycin–DNA Interaction†

Jonathan B. Chaires,* Keith R. Fox, Julio E. Herrera, Mark Britt, and Michael J. Waring
Department of Biochemistry, The University of Mississippi Medical Center, Jackson, Mississippi 39216-4505, and Department of Pharmacology, Medical School, University of Cambridge, Cambridge CB2 2QD, U.K.
Received February 24, 1987; Revised Manuscript Received August 10, 1987

ABSTRACT: The site and sequence specificity of the daunomycin–DNA interaction was examined by equilibrium binding methods, by deoxyribonuclease I footprinting studies, and by examination of the effect of the antibiotic on the cleavage of linearized pBR322 DNA by restriction endonucleases *Pvu*I and *Eco*RI. These three experimental approaches provide mutually consistent results showing that daunomycin indeed recognizes specific sites along the DNA lattice. The affinity of daunomycin toward natural DNA increases with increasing GC content. The quantitative results are most readily explained by binding models in which daunomycin interacts with sites containing two adjacent GC base pairs, possibly occurring as part of a triplet recognition sequence. Deoxyribonuclease I footprinting studies utilizing the 160 base pair (bp) *tyrT* DNA fragment and 61 and 53 bp restriction fragments isolated from pBR322 DNA further define the sequence specificity of daunomycin binding. Specific, reproducible protection patterns were obtained for each DNA fragment at 4 °C. Seven protected sequences, ranging in size from 4 to 14 bp, were identified within the *tyrT* fragment. Relative to the overall *tyrT* sequence, these protected sequences were GC rich and contained a more limited and distinct distribution of di- and trinucleotides. Within all of the protected sequences, a triplet containing adjacent GC base pairs flanked by an AT base pair could be found in one or more copies. Nowhere in the *tyrT* fragment did that triplet occur outside a protected sequence. The same triplet occurred within seven out of nine protected sequences observed in the fragments isolated from pBR322 DNA. In the two remaining cases, three contiguous GC base pairs were found. We conclude that the preferred daunomycin triplet binding site contains adjacent GC base pairs, of variable sequence, flanked by an AT base pair. This conclusion is consistent with the results of a recent theoretical study of daunomycin sequence specificity [Chen, K.-X., Gresh, N., & Pullman, B. (1985) *J. Biomol. Struct. Dyn. 3*, 445–466]. Adriamycin and the β-anomer of adriamycin produce the same qualitative pattern of protection as daunomycin with the *tyrT* fragment. Daunomycin inhibits the rate of digestion of pBR322 DNA by *Pvu*I (recognition sequence 5′-CGATCG-3′) to a greater extent than it does *Eco*RI (recognition sequence 5′-GAATTC-3′), a finding consistent with the conclusions derived from our footprinting studies. Our results, as a whole, are the clearest indication to date that daunomycin recognizes a specific DNA sequence as a preferred binding site.

The sequence specificity of antibiotic–DNA interactions is a topic of intense current interest (Gale et al., 1981; Neidle & Abraham, 1982; Dabrowiak, 1983; Dervan, 1986). The identification and characterization of preferred antibiotic binding sites within DNA sequences are essential for a detailed understanding of the molecular basis of antibiotic action. By understanding what specificity an antibiotic may have, and the molecular determinants of that specificity, it should be possible to develop a rational basis for the design of a new generation of chemotherapeutic agents of enhanced potency.

Daunomycin is an anthracycline antibiotic of clinical importance in cancer chemotherapy. Direct interaction with DNA is thought to be central to the molecular mechanism by which daunomycin acts (Arcamone, 1981). The drug intercalates into DNA and inhibits DNA replication and RNA transcription both in vivo and in vitro. After years of intensive study, daunomycin is now among the best characterized intercalators. The equilibrium (Schutz et al., 1979; Chaires et al., 1982; Graves & Krugh, 1983; Chaires, 1985a) and kinetic (Chaires et al., 1985) properties of the daunomycin–DNA interaction are well characterized. The structure of a daunomycin–oligonucleotide complex is known to atomic resolution (Quigley et al., 1980; Wang et al., 1987). The interaction of daunomycin with nucleosomes (Chaires et al., 1983) and with left-handed Z DNA (Chaires, 1983a, 1985b, 1986) has been explored. These studies show that the drug strongly prefers right-handed B-form DNA as a binding site and will selectively discriminate against alternate DNA conformations.

The possible site or sequence specificity of the daunomycin–DNA interaction remains, however, undefined. In contrast, other DNA binding antibiotics, notably actinomycin and netropsin, show pronounced site specificity (Neidle & Abraham, 1984; Dabrowiak, 1983). Solution studies utilizing synthetic deoxypolynucleotides attempting to elucidate the specificity of the daunomycin binding interaction have shown contradictory results [reviewed in Neidle and Sanderson (1983) and Chaires (1983b)]. Results from this laboratory using purified and well-characterized deoxypolynucleotides dem-

† Supported by U.S. Public Health Service Grant CA35635 (J.B.C.), by National Science Foundation Grant INT-8521004 (J.B.C.), and by grants from the Cancer Research Campaign, the Medical Research Council, and the Royal Society (M.J.W.).
* Address correspondence to this author at the Department of Biochemistry, The University of Mississippi Medical Center.

0006-2960/87/0426-8227$01.50/0