Sequence Requirements of the Hammerhead RNA Self-Cleavage Reaction†

Duane E. Ruffner,‡ Gary D. Stormo,§ and Olke C. Uhlenbeck*,‡

Departments of Chemistry and Biochemistry and of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309

Received June 13, 1990; Revised Manuscript Received August 7, 1990

ABSTRACT: A previously well-characterized hammerhead catalytic RNA consisting of a 24-nucleotide substrate and a 19-nucleotide ribozyme was used to perform an extensive mutagenesis study. The cleavage rates of 21 different substrate mutations and 24 different ribozyme mutations were determined. Only one of the three phylogenetically conserved base pairs but all nine of the conserved single-stranded residues in the central core are needed for self cleavage. In most cases the mutations did not alter the ability of the hammerhead to assemble into a bimolecular complex. In the few cases where mutant hammerheads did not assemble, it appeared to be the result of the mutation stabilizing an alternate substrate or ribozyme secondary structure. All combinations of mutant substrate and mutant ribozyme were less active than the corresponding single mutations, suggesting that the hammerhead contains few, if any, replaceable tertiary interactions as are found in tRNA. The refined consensus hammerhead resulting from this work was used to identify potential hammerheads present in a variety of *Escherichia coli* gene sequences.

The RNA genomes of several virusoids contain a domain of approximately 50 nucleotides that is required for an autocatalytic cleavage step in their replication pathway (Forster & Symons, 1987a). A "hammerhead" secondary structure composed of three helices joined at a central core of 11–12 single-stranded nucleotides has been identified as necessary and sufficient for the self-cleavage reaction (Uhlenbeck, 1987; Forster & Symons, 1987b). The helices are stable enough to permit the assembly of active hammerheads from two or even three separate RNA molecules (Uhlenbeck, 1987; Haseloff & Gerlach, 1988; Koizumi et al., 1988; Jeffries & Symons, 1989), and catalytic cleavage can be demonstrated through successive cycles of helix annealing and dissociation (Uhlenbeck, 1987; Haseloff & Gerlach, 1988; Jeffries & Symons, 1989; Koizumi et al., 1989). This system is the smallest example of a catalytic RNA found to date.

It is clear that the hammerhead must have a fairly precise sequence requirement for cleavage. All three helices and thirteen conserved nucleotides appear in the seven available natural examples of hammerheads (Forster & Symons, 1987a,b). A limited number of mutagenesis experiments have confirmed the need for complementary base pairs in the helices and for specific nucleotides at several of the single-stranded positions (Sampson et al., 1987; Koizumi et al., 1988, 1989; Sheldon & Symons, 1989; Ruffner et al., 1989). Presumably the essential nucleotides could either participate in tertiary interactions, coordinate catalytically important divalent metal ions, or provide functional groups necessary for activation of the labile bond. The intent of this paper is to systematically determine the sequence requirements of the hammerhead domain by changing each of the nucleotides in the conserved or semiconserved positions of the hammerhead to each of the other nucleotides and examining their activity. The resulting data allow for the revision of the hammerhead consensus sequence, and this revised consensus was used to search other RNA sequences for hammerheads. Although it is difficult to use mutagenesis experiments to learn about structure, some insight into the structure of the hammerhead was obtained.

Materials and Methods

RNA Synthesis. RNA was synthesized by transcription with T7 RNA polymerase, of oligodeoxyribonucleotide templates, as previously described (Milligan et al., 1987). Transcription reactions contained 100 nM template, 0.1 mg/mL T7 RNA polymerase, 2 mM of each of the 4 NTPs,[1] and 12 mM $MgCl_2$. The transcripts were internally labeled by the inclusion in the transcription reaction of one of the [α-^{32}P]NTPs to give a specific activity of ~0.5 Ci/μmol of RNA. For determining the concentration of unlabeled transcripts a residue extinction coefficient of 8.3×10^3 M^{-1} cm^{-1} at 260 nm was assumed.

Full-length transcripts were separated from unincorporated NTPs and abortive transcription products by gel electrophoresis in denaturing polyacrylamide gels. The transcripts were localized in the gel either by autoradiography or by UV shadowing and subsequently extracted by a crush and soak method. The transcripts were concentrated by ethanol precipitation. The sequences of several of the trascripts were checked by standard enzymatic (Donis-Keller et al., 1977; Simoncsits et al., 1977) or chemical sequencing methods (Peattie, 1979) or, alternatively, by using a new method based on phosphorothioate chemistry (Gish & Eckstein, 1988). The remaining transcripts were identified by their comigration on denaturing acrylamide gels with the sequenced transcripts.

Cleavage Reactions. Cleavage reactions contained 0.2 μM substrate, 0.3 μM enzyme, 50 mM Tris-HCl (pH 8), and 10 mM $MgCl_2$, except where otherwise indicated. The enzyme and substrate were combined in the presence of buffer and incubated at 60 °C for 1 min. The reactions were started by adding $MgCl_2$, to a final concentration of 10 mM, at 37 °C. Aliquots were removed from the reactions at various times and stopped by the addition of 1 volume of 50 mM EDTA, 7 M urea, and 0.05% xylene cyanol and bromophenol blue dyes. For rate determinations five time points were taken ranging

†G.D.S. was supported by NIH Grants GM28755 and GM41776. D.E.R. and O.C.U. were supported by NIH Grants GM36944 and AI30242.

* Address correspondence to this author.

‡ Department of Chemistry and Biochemistry.

§ Department of Molecular, Cellular and Developmental Biology.

[1] Abbreviations: NTP, any nucleoside triphosphate; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid

0006-2960/90/0429-10695$02.50/0

from approximately 0 to 60% total cleavage. Uncleaved substrate and products were separated by electrophoresis in denaturing polyacrylamide gels, and extents of cleavage were determined by quantitation of radioactivity in the substrate and product bands either by using an Ambis gel scanner or by Cerenkov counting of the respective bands. The logarithm of the uncleaved fraction was plotted versus time, and the data points were fitted to a line by a least-squares method. The cleavage half-lives ($t_{1/2}$) derived from the least-squares fit were used to obtain first-order rate constants by using the relationship $k = 0.693/t_{1/2}$.

Native Gel Electrophoresis. Labeled wild-type S was combined with nonlabeled wild-type or mutant E, and labeled wild-type E was used with the wild-type or mutant S. Labeled and nonlabeled RNAs were present at 0.1 and 0.2 μM, respectively. Tris-HCl (pH 8) was added to a final concentration of 50 mM. The mixture was incubated at 65 °C for 0.5 min and 37 °C for 2 min and subsequently placed on ice. One volume of native gel dye (40% sucrose, 0.5% xylene cyanol and bromophenol blue dyes) was added, and the samples were electrophoressed in 15% (19:1) acrylamide gels containing 50 mM Tris–HOAc (pH 8) and 10 mM $Mg(OAc)_2$. Electrophoresis was performed at 4 °C.

Search for Hammerheads in Escherichia coli Sequences. All *E. coli* sequences in Genbank release 48 were examined for the presence of potential hammerhead sequences compatible with the revised hammerhead consensus (Figure 6A). For the purposes of the search the hammerhead domain was divided into three segments (a, b, and c; Figure 6A) whereby each participates in Watson–Crick base pairing with the two other segments. The search was performed in two parts. First, all *E. coli* sequences in Genbank were searched for the presence of any of the 64 combinations of strands a and b, where a and b are defined as XXXctgangarNNN and N′N′N′ygaaaZZ, respectively. For this part of the search X and Z were not specified and, y and r must be C or T and A or G, respectively. A match was scored only when a and b were both present in a single Genbank entry and N′N′N′ and NNN were reverse complements of each other. This portion of the search was performed by using the Fast Data Finder (FDF) System at Applied Biosystems (Yu et al., 1987).

In the second part of the search the Genbank entries which contained matches were retrieved and examined for the presence of the third (c) and final portion of the hammerhead by using the pattern searching routine of the EuGene programs (Baylor College of Medicine). The retrieved entries were searched for matches to the following pattern: Z′Z′uHX′X′X′, where H is any nucleotide except G and Z′Z′ and X′X′X′ are the reverse complements of ZZ and XXX of the a and b sequences identified in the first part of the search. The order of, or the distance between, parts a, b, and c was not constrained during either part of the search.

Results

A previously well-characterized hammerhead consisting of two synthetic oligoribonucleotides was used as a basis for mutagenesis experiments (Uhlenbeck, 1987). As depicted in Figure 1A, a 24-nucleotide substrate (S) pairs through helices I and II to a 19-nucleotide enzyme (E). The resulting hammerhead is consistent with the proposed consensus structure (Forster & Symons, 1987a,b) in Figure 1B. Cleavage occurs rapidly to give an 18-nucleotide P1 and a 6-nucleotide P2. Under conditions of substrate excess, the catalytic cleavage of S was found to have $K_m = 0.6$ μM and $k_{cat} = 0.5$/min at 55 °C in a standard buffer. Unlike several other hammerheads (Forster et al., 1988), this pair of oligonucleotides cleaves as a two-stranded complex rather than a four-stranded "double hammerhead" (Ruffner et al., 1989).

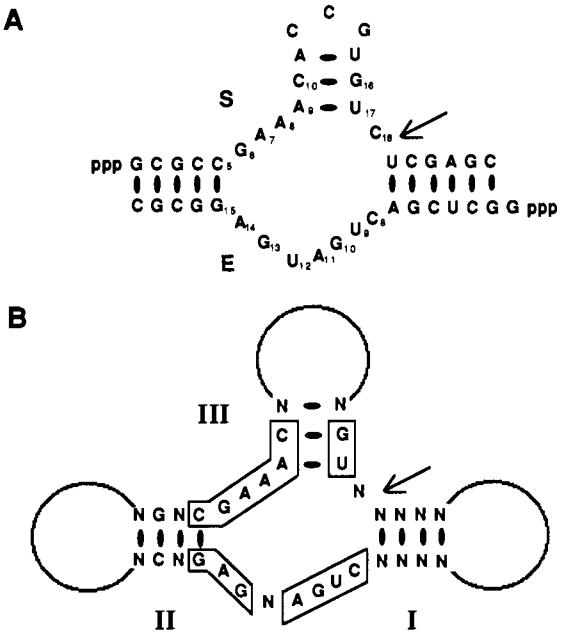

FIGURE 1: (A) A synthetic hammerhead composed of a 24-nucleotide substrate (S) and a 19-nucleotide enzyme (E) was used as a basis for mutagenesis. The arrow indicates the site of cleavage, as previously determined (Uhlenbeck, 1987). The nucleotides numbered in both the substrate and enzyme are those positions that were mutagenized. The numbers indicate the relative position of each nucleotide from the 5′ end of their respective strands. Nucleotide substitutions will be identified by naming the wild-type nucleotide followed by the position number and the nucleotide substitution. For example, substitution of C for the A at position 7 of the substrate will be designated S-A7C. Base pair mutants will be designated similarly except that the identity of the two wild-type nucleotides will be dropped and a "/" will separate the two base-paired nucleotide substitutions. For example, substitution of a U/A base pair in place of the wild-type A/U pair formed between positions 9 and 17 of the substrate will produce a mutant substrate designated S-U9/A17. (B) A hammerhead consensus structure. The hammerhead domain is composed of a conserved central core, shown single stranded, held together by three stems designated I, II, and III. Watson–Crick base pairs in the stems are indicated with ellipses. Boxed positions are those previously believed to be essential due to their conservation in seven naturally occurring hammerhead domains (Forster & Symons, 1987a,b).

The residues that were changed are indicated by their residue numbers in Figure 1A. The nine nucleotides in S and the eight nucleotides in E include all the single-stranded residues in the central core of the domain as well as three conserved base pairs identified by Forster and Symons (1987a,b). Mutant hammerheads were tested for cleavage at 0.3 μM E and 0.2 μM S at 37 °C in 10 mM $MgCl_2$ and 50 mM Tris, pH 8.0. Under these conditions, the time required for 50% cleavage is a convenient 5 min for the wild-type sequence. By using a slight excess of enzyme over substrate, effects of product release on reaction rate are avoided. In a thorough study of a series of hammerheads with similar helix lengths, cleavage chemistry rather than product release was believed to be rate limiting (Fedor & Uhlenbeck, 1990). Under the reaction conditions used, the concentrations of oligonucleotides are substurating so that mutations which affect either the stability of the complex or the catalytic rate will be identified. The 37 °C reaction temperature was chosen to be well below the optimal value of 55 °C to ensure that potentially less stable mutants could still assemble.

The two inner base pairs of stem III, S-A9/U17 and S-C10/G16, were each changed to every other pair to give six

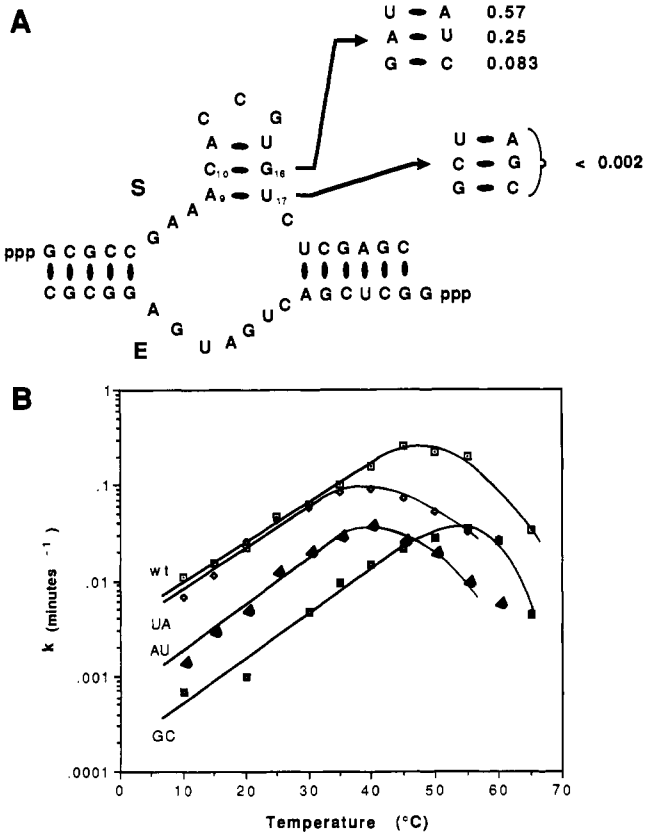

FIGURE 2: Mutagenesis of the conserved base pairs of stem III. (A) Each base pair was replaced with all three other Watson–Crick pairs, and the rates of cleavage were examined. The cleavage rates were normalized to the wild-type E–S pair, and the numbers indicate the relative rates of cleavage (k_{rel}). (B) Temperature dependence of cleavage of the wild-type and the N10/N16 base pair mutants. First-order rate constants for wild-type or N10/N16 base pair mutants were determined, at temperatures from 10 to 65 °C.

variant S molecules. When combined with E, any substitution of the A9/U17 pair was found to result in virtually complete loss of activity (Figure 2A). In contrast, all mutants of the C10/G16 pair were active, but the rates differed. The U10/A16 variant showed virtually identical activity, while the two R10/Y16 pairs cleaved at substantially lower rates. Examination of the temperature dependence of the cleavage rate for the N10/N16 variants revealed a similar slope at lower temperatures, but different optimal temperatures (Figure 2B). Presumably the lower temperature optima for the A10/U16 or U10/A16 substitutions is due to the lower stability of stem III resulting from replacement of the wild-type C/G base pair. The overall lower cleavage rates for both variants containing a purine (R) at position 10 and pyrimidine (Y) at position 16 (i.e., A10/U16 and G10/C16) are less easily explained but are probably not due to the effects on the stability of stem III. Perhaps both R10/Y16 pairs place the essential A9/U17 pair in a less favorable orientation due to differences in stacking on a R/Y versus a Y/R pair.

The inner base pair of stem II, formed between S-C5 and E-G15, was altered to every possible combination of nucleotides (Figure 3). Although the wild-type S-C5/E-G15 pair was the most active, four of the eight substrate/enzyme combinations containing a pyrimidine at position 5 of the substrate cleaved at a rates within 30% of the wild-type rate. Even combinations containing mismatches at this position are active when a pyrimidine is present in the substrate strand. In contrast, hammerheads containing a purine at position 5 of the substrate show much lower rates of cleavage whether or not they are paired. This rather complex sequence requirement

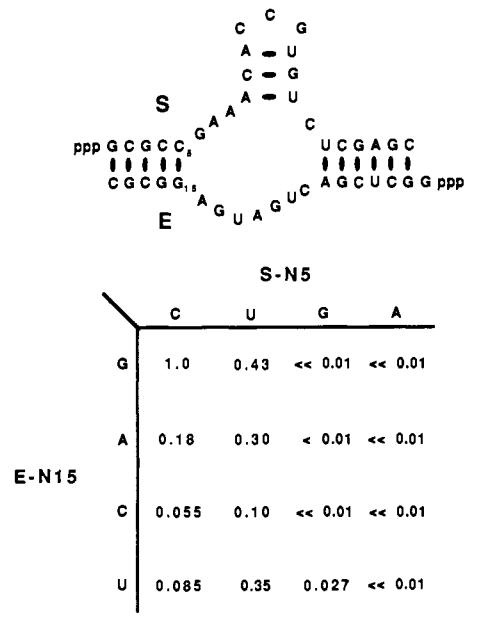

FIGURE 3: Mutagenesis of the conserved base pair of stem II. Positions 5 and 15 of the substrate and enzyme, respectively, were replaced with all three other nucleotides. All combinations of E-G15N and S-C5N were combined and the cleavage rates normalized to that of the wild-type enzyme–substrate pair.

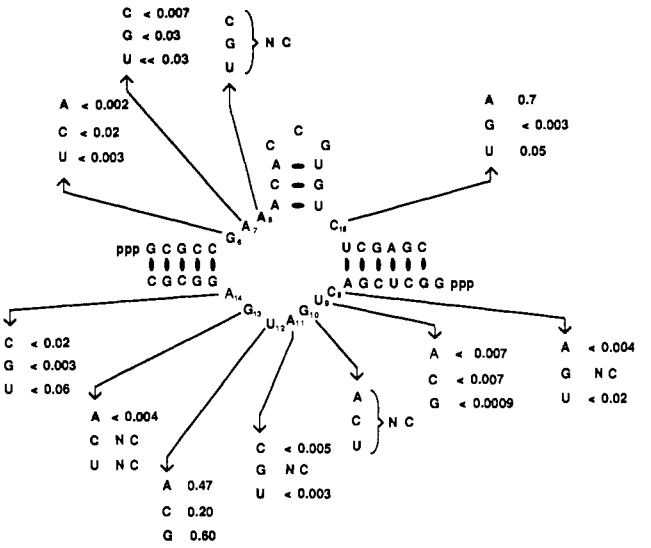

FIGURE 4: Mutagenesis of nucleotides in the central core. All nucleotides depicted as single stranded in Figure 1A were replaced separately with all three other nucleotides. Mutant enzyme or substrate was combined with their respective wild-type substrate or enzyme and examined for cleavage activity. The relative cleavage rates, with respect to the wild-type E–S pair, are indicated. NC indicates that no cleavage was detected after 2 h of incubation under standard conditions.

suggests that the nucleotides at positions S-5 and E-15 contribute to hammerhead structure or catalysis in a manner more complex than by forming a simple base pair.

Mutations in each of the 11 single-stranded residues in the hammerhead core were tested by combining a mutant E or S with corresponding wild-type substrate or enzyme. The cleavage rates of all 33 mutants were less than wild type, although the magnitude of the effect varied considerably depending on both the position substituted and the type of substitution (Figure 4). All three mutants of E-U12 were quite active hammerheads, which is consistent with the fact that three of the four nucleotides are present in natural hammerhead sequences. The nucleotide 5′ to the cleavage site, S-C18, shows an interesting sequence dependence. An A at

that position cleaves nearly as well as a C, but a U cleaved considerably more slowly and a G hardly cleaved at all. Of the nine remaining single-stranded residues, mutants of seven showed slow, but detectable, cleavage with rates that varied from 0.06 to 0.001 of the wild-type rate. However, any mutation in the two remaining residues, S-A8 and E-G10, showed no trace of cleavage under standard cleavage conditions.

Since the deleterious effects of mutations in tRNA (Potts et al., 1981) and the self-splicing intron of *Tetrahymena* rRNA (Zaug et al., 1988) can sometimes be partially restored by altering the reaction conditions, several hammerhead mutants were examined under different reaction conditions. For 19 E variants (E-N8 through N11 and E-N13 through N14) and four S variants (S-N8), cleavage reactions were carried out in the standard buffer with a variety of additions [including 5 mM spermidine, 60 mM $MgCl_2$, 0.5 M NaCl, 10% (v/v) acetone, acetonitrile, or ethanol] and at varying temperatures from 15 to 50 °C. In no instance was the rate of cleavage of the mutant increased relative to wild type. Significantly, no cleavage was observed for any S-N8 mutant after 4 h of incubation in the standard reaction buffer at 24, 37, and 50 °C. However, trace amounts of cleavage were observed for S-A8U after 4 h of incubation with the addition of Mg to a final concentration of 70 mM or the inclusion at 10% of any of the three organic solvents examined.

One possible role for the many essential nucleotides in the hammerhead is to maintain the tertiary structure of the molecule. If this is the case, a mutation could lower the affinity of E with S, thereby reducing the cleavage rate. In order to monitor the binding of E to S, mixtures of oligomers were analyzed on nondenaturing polyacrylamide gels (Figure 5). ^{32}P-Labeled wild-type substrate was combined with unlabeled enzyme variants, or ^{32}P-labeled wild-type enzyme was combined with unlabeled substrate variants. At concentrations comparable to those in the cleavage reactions, about 50% of the wild-type E–S pair forms a slower migrating complex and the rest remains free (labeled "wt" in Figure 5A,B). The complex was eluted from the gel and, upon subsequent analysis on denaturing gels, was found to contain P1 and E. Cleavage had occurred before or during electrophoresis, and P2 had dissociated from the cleaved complex.

The remainder of the lanes in Figure 5 show complexes formed between ^{32}P-labeled E and nonradioactive mutant substrates (panel A) or ^{32}P-labeled S and nonradioactive mutant enzymes (panel B). Depending upon the mutation, three different types of results were obtained. In most of the cases, about half of the radioactivity migrated at a position corresponding to an uncleaved bimolecular complex. Since this is about the same proportion of complex that forms with the wild-type hammerhead at the same oligomer concentration, it appears that the mutations do not greatly alter the ability of the two oligomers to interact and therefore the poor cleavage is not the result of their inability to assemble. In four cases (S-A8G, S-A10/U16, S-U9/A17, and S-C18U) very little bimolecular complex is formed. For two of these four mutations, other mutations at the same position do not prevent assembly and also do not cleave. The poor assembly of mutants of these positions is therefore unlikely to be the cause of their reduced rate of cleavage. However, the poor cleavage of the S-C18U mutation may be the result of poor assembly since S-C18A cleaves well and other hammerheads with a U5′ to the cleavage site have been found to be active. Finally, in nine cases (S-G9/C17, S-U10/A16, E-U9C, E-A11N, E-U12G, E-G13A, and E-A14C) the mutation appeared to result in more bimolecular complex than wild type.

Interpretation of the results in Figure 5 is complicated by the fact that both the E and S oligoribonucleotides have a secondary structure of their own which must be at least partially disrupted to form the hammerhead (Ruffner et al., 1989). In the case of S, this structure is likely to be a hairpin which continues helix III for at least three base pairs after a A8/C18 mismatch (Figure 5C). In the case of E, it has not been possible to propose a structure involving only Watson–Crick base pairs.

The anomalous assembly of many of the mutant substrate oligoribonucleotides can be explained in terms of alternate structures. The S-A8G and S-C18U mutations change the A8/C18 mismatch in the hairpin into a normal base pair (Figure 5C), resulting in a more stable structure that will be harder to disrupt to form the hammerhead. Similarly, a different, possibly more stable structure for S-A10/U16 can be proposed (Figure 5D). However, the hairpin structure cannot explain the unusual pattern of results seen for base pair substitutions of the A9/U17 pair of S. While all three substitutions are inactive, one inhibits assembly, another improves it, and a third has no effect. Since A9/U17 is a base pair in both the hammerhead and the hairpin of S, it is not clear whey these mutations would affect assembly.

It is reasonable to also explain the anomalous assembly of mutant E oligomers in terms of alternate structures. However, the absence of a clear structure for E makes interpretation difficult.

Considering the large number of essential single-stranded nucleotides in the hammerhead, it is possible that a number of these residues are forming specific tertiary interactions. These could involve standard Watson–Crick pairs, a variety of possible non-Watson–Crick pairs, or base triples. Since many of the tertiary interactions in tRNA can be replaced by different nucleotides that maintain pairing without altering the backbone (Sampson et al., 1987, 1990; Behlen et al., 1990), it seemed worthwhile to search for tertiary interactions in the hammerhead. Toward this end, we attempted to restore activity to a mutant E or S by combining it with a second mutant S or E. Thus, 12 mutants of S (three each at positions 6, 7, 8, and 9/17) were individually combined with 18 mutants of E (three each at positions 8, 9, 10, 11, 13, and 14) under standard reaction conditions. Additionally, the 3 variants of S-C18N were combined with the 3 variants of E-C8N, and S-C18G was combined with 18 of the variants of E (three each at positions 8, 9, 10, 11, 13, and 14). In every one of the 243 reactions, the cleavage rate for the hammerhead containing two mutations was less than that of the corresponding single mutants (data not shown). Thus no evidence for tertiary pairs was obtained among the double mutants tested.

The results of the experiments with two mutant oligoribonucleotides suggests a reasonable explanation for the poor cleavage of S-C18G, the hammerhead with a G residue 5′ to the cleavage site. As reported in Figure 4, S-C18A is fully active with the wild type E and a slow, but measurable, rate of cleavage was observed when E-C8U is combined with the wild-type S. However, when S-C18A is combined with E-C8U, the cleavage rate was much lower than with the E-C8U mutant alone. This result can be explained by proposing that the double mutant extends helix I by forming a base pair between S-A18 and E-U8 that is incompatible with an active hammerhead conformation. Therefore, the poor cleavage of S-C18G with the wild-type E could also be the result of the formation of a base pair between S-G18 and E-C8.

The mutation data presented in this work permits the consensus hammerhead structure of Forster and Symons to

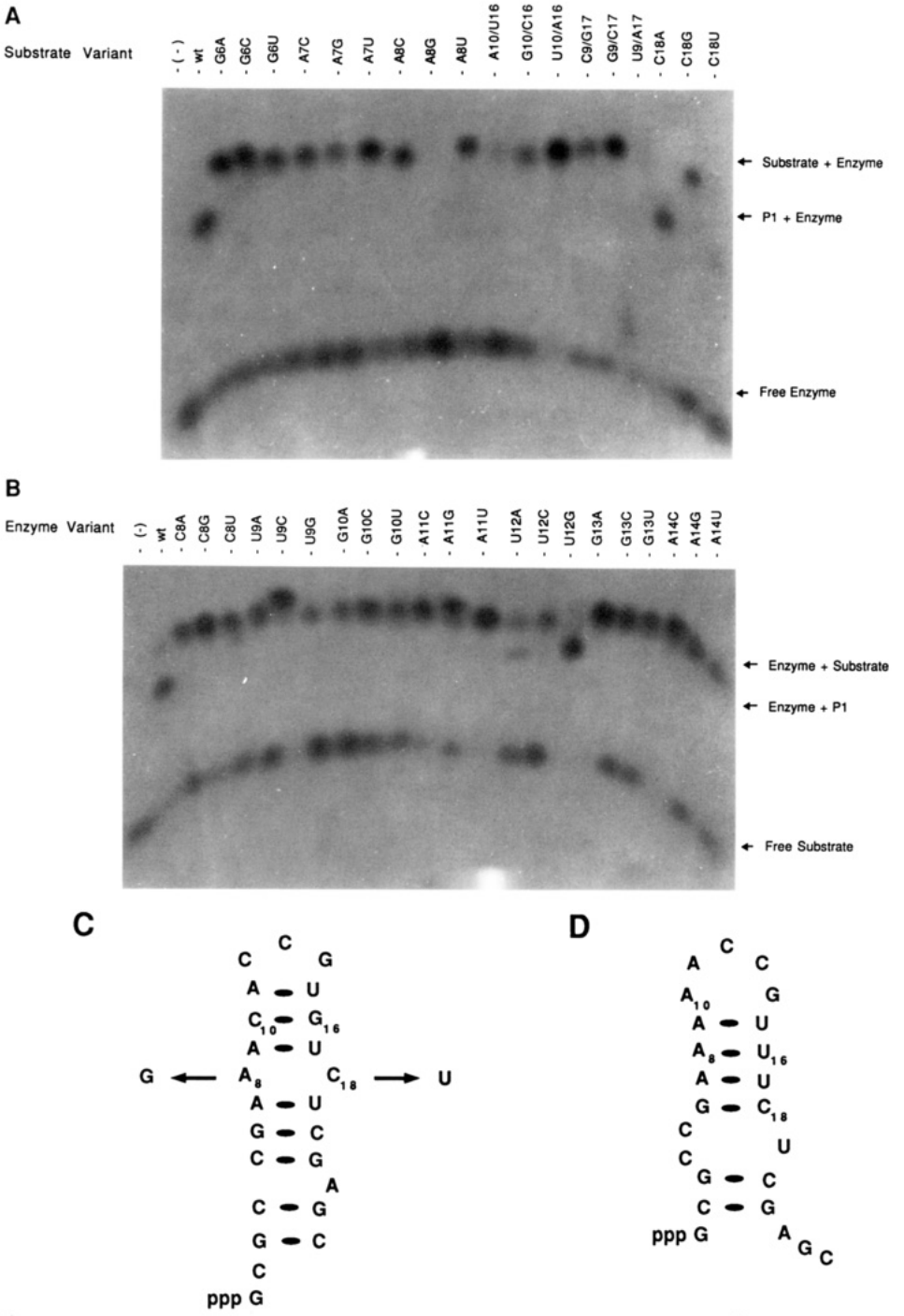

FIGURE 5: (A and B) Native gel electrophoresis. Substrate and enzyme mutants were combined with ^{32}P-labeled E (panel A) or ^{32}P-labeled S (panel B), respectively, and examined by native gel electrophoresis for their ability to form E–S complex. (C) A potential hairpin structure of S indicating the A8G and C18U mutations which could further stabilize this structure. (D) A different competing hairpin structure for S-A10/U16.

be refined. Since E-C12 is the slowest cleaving mutation at a nonconserved position (aside from S-U18 which is artifactually slow) and a naturally occurring hammerhead contains a C at the corresponding position (vSCMoV; Keese & Symons, 1987), then any mutation with $k_{rel} \geq 0.2$ should be considered active. Futhermore, the conserved base pair corresponding to S-C10/G16 appears to be conserved for reasons other than its effect on cleavage efficiency since all base pair substitutions are active and the least active is just under 10-fold slower than the wild type. Therefore, the minimal cutoff for active hammerheads should be mutants with a $k_{rel} \approx 0.1$ or greater. On the basis of these considerations, a revised hammerhead consensus is depicted in Figure 6A. The features of the consensus include the following: there are 11 conserved nucleotides (including the lower conserved base pair of stem III); any nucleotide except G is allowed 5′ of the cleavage site; and the innermost base pair of stem II is designated as a Y/R pair, but actually has a more complex sequence requirement (see Figure 3). In addition to these requirements a minimum length for each of the three helices is assumed to be three Watson–Crick base pairs although this was not empirically tested in this study. It should be noted that the consensus does not take into account the possibility that extra nucleotides can be inserted into various positions in the hammerhead and that G/U base pairs may be tolerated in the helices. While this issue remains to be systematically explored, hammerheads with an

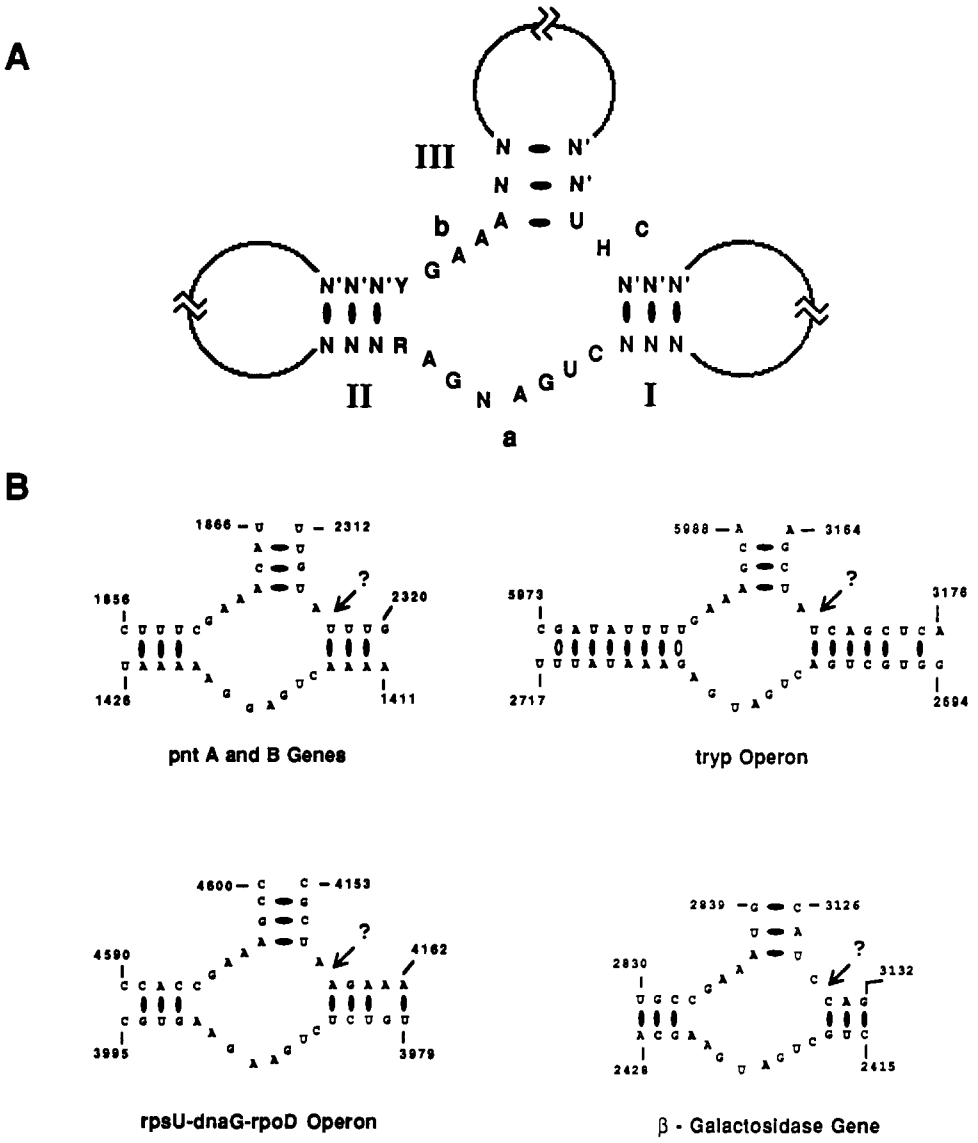

FIGURE 6: (A) Revised hammerhead consensus used in the data base search. Y, R, N, and H designate either pyrimidine, either purine, any nucleotide, and any nucleotide except G, respectively. N′ designates the Waston–Crick base pair of the corresponding N. On the basis of the results presented in Figure 3, the sequence requirements of the Y/R pair are actually more complicated but were not used in the search. Ellipses designate base pairs. (B) Potential hammerhead domains in *E. coli*. The hammerhead domains, starting with the upper left and continuing clockwise, were located in the following *E. coli* sequences as identified by their Genbank loci: ECOPNTAB, ECOTGP, ECOLAC, and ECORPSRPO. The nucleotide position numbers are as reported in Genbank. Putative cleavage sites are indicated by arrows.

additional U or C between E-14 and 15 and between E-7 and 8 are active (Forster & Symons, 1987b; Sheldon & Symons, 1989) and a hammerhead containing a G/U base pair corresponding to S-C10/G16 is active (Sheldon & Symons, 1989). Additionally, several variants of the avocado sun blotch viroid have been identified which contain base pair mismatches in stem II (Rakowski & Symons, 1989).

If one analyzes the consensus as it appears in Figure 6 for its information content (Schneider et al., 1986), the pattern contains 40.4 bits of information. However the expected frequency of potential hammerheads is complicated by the variable lengths of the loops connecting the conserved regions. In, fact if one assumes that the three parts of the pattern must merely be on the same RNA and may occur in any order at any distance from one another, the expected frequency of such patterns depends on the cube of the RNA length. For an RNA of length L, the expected number of potential hammerheads is $(L^3)(2^{-40.4})$. This expected number equals one at lengths of about 11 000 nucleotides. Thus, one expects to find hammerheads fairly often in RNA sequences, especially in long RNAs. A search of all *E. coli* sequences in Genbank release 48 identified many such potential hammerheads. Four of the more plausible ones are shown in Figure 6B.

Discussion

A systematic substitution of nucleotides in a model hammerhead self-cleaving domain reveals that the nucleotides in no less than 13 positions must be specified in order to obtain a reasonable rate of cleavage. Since so many residues are important, the resulting consensus hammerhead structure is quite similar to the one deduced by Forster and Symons which is based on relatively few natural examples. The results presented here are generally consistent with more limited mutagenesis studies with several other hammerheads. Using a hammerhead sequence derived from the Newt constructed from two molecules, Ohtsuka and colleagues (Koizumi et al., 1988) concentrated efforts on residues corresponding to E-C8 and S-C18, and the A9/U17 and C10/G16 pairs in S. Since no kinetic data were obtained and the Newt sequence is capable of forming a double hammerhead (Forster et al., 1988), comparison with the data presented here is difficult. However, their data confirm that G is not allowed at positions E-8 or

S-18, C ≅ U ≅ A at position S-18, and base pair changes of C10/G16 are active and A9/U17 are not. In addition, by showing that mismatches in the base pair corresponding to S-C10/G16 and S-A9/U17 produce inactive hammerheads, they established the requirement for a base pair at these positions (Koizumi et al., 1988, 1989). Sheldon and Symons prepared a number of mutants of the hammerhead from the plus strand of sLTSV (Sheldon & Symons, 1989). Since the RNA was made as a single transcript, it could cleave in the transcription reaction, preventing kinetic analysis. They showed that G/U pairs were active at A9/U17 and C10/G16 and additional nucleotides could be accommodated at certain positions in the molecule.

The sequence-specific cleavage of yeast $tRNA^{Phe}$ by lead is often considered analogous to hammerhead cleavage (Brown et al., 1983). Although this reaction probably occurs through a different mechanism, the available crystal structure of $tRNA^{Phe}$ provides considerable insight into the relation of RNA structure to the cleavage mechanism. In the region of intersection between the T and D loops, the reactive $Pb(OH)^+$ is coordinated to functional groups of two single-stranded pyrimidines and in such a way that it can remove the proton from the 2′-hydroxyl of ribose 17, resulting in cleavage in the D loop. A study examining the rate of lead cleavage for a variety of $tRNA^{Phe}$ mutants emphasize the importance of tertiary structure in the reaction (Behlen et al., 1990). Mutations that disrupt the secondary and tertiary interactions involved in maintaining the lead binding pocket result in a reduction in cleavage rate even when the changes occur very far from where the lead binds. On the other hand, mutations in single-stranded residues not involved in lead binding or tRNA folding have no effect on the reaction. Finally, in several cases, it is possible to make multiple changes that maintain tertiary interactions without altering cleavage. It would be reasonable to expect that corresponding observations would be seen in mutagenesis of the hammerhead.

Although the mechanism of hammerhead cleavage remains unknown, one can identify several features that will clearly require a precisely folded RNA structure. An extensive study on the divalent ion requirement of the reaction suggests the presence of a tightly bound "catalytic" divalent that is coordinated to one of the axial oxygens of the phosphate at the cleavage site and may be involved in transition-state stabilization (S. C. Dahm and O. C. Uhlenbeck, to be submitted for publication). In addition, it is possible that one or more nucleotide functional groups provide acid or base functions generally believed to be essential for RNA cleavage. Presumably the essential nucleotides in the hammerhead participate in tertiary interactions needed to maintain the structure as well as provide functional groups for the ion binding pocket and the catalytic mechanism. As in the lead cleavage domain, mutations can have little effect, greatly reduce, but not eliminate, the reaction, or totally abolish cleavage. In this regard, attention should be focused on S-A8 as a residue that may have a crucial role in the mechansim.

While not all possible multiple substitutions have been tested, thus far no combination of any two separate substitutions has resulted in efficient cleavage. This suggests that the hammerhead, unlike the lead cleavage domain, either does not have any tertiary interactions or, more likely, that existing tertiary interactions cannot be substituted with other nucleotides. It is possible, for example, that tertiary interactions are present but one or more functional groups of the participating nucleotides also provide a second function (such as metal binding) that cannot be replaced. It is known that certain tertiary interactions in the lead cleavage domain cannot be substituted in any way without disrupting the structure and hence cleavage. A striking example is the conserved Watson–Crick base pair G53/C61, where the amino group of C makes a H bond across the T loop to phosphate 60. Mutation of this pair to C/G drastically reduces cleavage by lead (Behlen et al., 1990).

The sequence requirements for hammerhead cleavage are sufficiently simple that many RNA molecules contain the information needed to undergo self-cleavage. However, since RNA can adopt many different conformations, the presence of a hammerhead sequence does not necessarily mean that cleavage will occur. For example, ASBV RNA uses hammerhead cleavage in its replication cycle but exists as a circular monomer (Palukaitis et al., 1979). Presumably, the hammerhead forms transiently during the replication cycle, but in the circular monomer it is in an alternate inactive conformation. Even when a hammerhead is in a contiguous RNA sequence, cleavage does not necessarily occur during in vitro transcription and a denaturation/renaturation cycle is required to cause cleavage (Forster & Symons, 1987a). Thus the presence of potential hammerheads in a variety of bacterial mRNAs is of unknown significance. Cleavage may only occur in particular circumstances or never at all. However, it is clear that the potential of many RNAs to undergo self-cleavage has implications on RNA stability in vivo.

ACKNOWLEDGMENTS

We thank Elaine Mansfield, Carlos Zamudio, and Ken Takara from Applied Biosystems for the FDF search.

REFERENCES

Behlen, L. S., Sampson, J. R., Direnzo, A. B., & Uhlenbeck, O. C. (1990) *Biochemistry 29*, 2515–2523.

Brown, R. S., Hingerty, B. E., Dewan, J. C., & Klug, A. (1983) *Nature 309*, 543–546.

Donis-Keller, H., Maxam, A. M., & Gilbert, W. (1977) *Nucleic Acids Res. 4*, 2527–2538.

Epstein, L. M., & Gall, J. G. (1987) *Cell 48*, 535–543.

Fedor, M., & Uhlenbeck, O. C. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1668–1672.

Forster, A. C., & Symons, R. H. (1987a) *Cell 49*, 211–220.

Forster, A. C., & Symons, R. H. (1987b) *Cell 50*, 9–16.

Forster, A. C., Davies, C., Sheldon, C. C., Jeffries, A. C., & Symons, R. H. (1988) *Nature 334*, 265–267.

Gish, G., & Eckstein, F. (1988) *Science 240*, 1520–1522.

Haseloff, J., & Gerlach, W. L. (1988) *Nature 334*, 585–591.

Jeffries, A. C., & Symons, R. H. (1989) *Nucleic Acids Res. 17*, 1371–1377.

Keese, P., & Symons, R. H. (1987) The Structure of Viroids and Virusoids. In *Viroids and Viroid-like Pathogens* (Semancik, J. S., Ed.) CRC Press, Boca Raton, FL.

Koizumi, M., Iwai, S., & Ohtsuka, E. (1988) *FEBS Lett. 239*, 285–288.

Koizumi, M., Hayase, Y., Iwai, S., Kamiya, H., Inoue, H., & Ohtsuka, E. (1989) *Nucleic Acids Res. 17*, 7059–7070.

Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.

Paulkaitis, P., Hatta, T., Alexander, D. M., & Symons, R. H. (1979) *Virology 99*, 145–151.

Peattie, D. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1760–1764.

Potts, R., Ford, N., & Fournier, M. (1981) *Biochemistry 20*, 1653–1659.

Rakowski, A. G., & Symons, R. H. (1989) *Virology 173*, 352–356.

Ruffner, D. E., Dahm, S. C., & Uhlenbeck, O. C. (1989) *Gene 82*, 31–41.
Sampson, J. R., Sullivan, F. X., Behlen, L. S., Direnzo, A. B., & Uhlenbeck, O. C. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 267–275.
Sampson, J. R., Direnzo, A. B., Behlen, L. S., & Uhlenbeck, O. C. (1990) *Biochemistry 29*, 2523–2532.
Schneider, T. D., Stormo, G. D., Gold, L., & Ehrenfeucht, A. (1986) *J. Mol. Biol. 188*, 415–431.
Sheldon, C. C., & Symons, R. H. (1989) *Nucleic Acids Res. 17*, 5679–5685.
Simoncsits, A., Brownlee, G. G., Brown, R. S., Rubin, J. R., & Guilley, H. (1977) *Nature 269*, 833–836.
Uhlenbeck, O. C. (1987) *Nature 328*, 596–600.
Yu, K.-I., Hsu, S.-P., Heiss, R. E., Jr., & Hasiuk, L. Z. (1987) *Quest 4*, 5–19.
Zaug, A. J., Grosshans, C. A., & Cech, T. R. (1988) *Biochemistry 27*, 8924–8931.

Heterogeneous Initiation Due to Slippage at the Bacteriophage 82 Late Gene Promoter in Vitro†

Hwai-Chen Guo‡ and Jeffrey W. Roberts*
Section of Biochemistry, Molecular and Cell Biology, Biotechnology Building, Cornell University, Ithaca, New York 14853
Received May 4, 1990; Revised Manuscript Received August 6, 1990

ABSTRACT: RNAs synthesized in vitro by purified *Escherichia coli* RNA polymerase from a bacteriophage 82 promoter are heterogeneous at the 5′ end. We show that this heterogeneity results from variable addition of extra adenine residues, allowed by slippage of the initial oligonucleotide pppAAA-OH against its DNA template sequence TTT. Slippage backward by one base allows another A to be added, giving pppAAAA-OH, and this cycle can continue more than 20 times before it is ended by incorporation of UMP encoded by the fourth template base A. Slippage is abolished by mutation of the TTT template sequence to TGT and is sensitive to the concentrations of UTP and ATP in the reaction mixture. Analysis of deletions, substitutions, and point mutants implies that the slippage reaction requires only the existence of TTT at the initiation site of the template strand, although changes in neighboring nucleotides slightly affect its efficiency.

Initiation of an RNA chain by *Escherichia coli* RNA polymerase is thought to comprise a defined sequence of steps: (i) RNA polymerase locates a promoter and binds specifically to form an inactive intermediate, the "closed complex"; (ii) the closed complex isomerizes into an "open complex" in which about 16 nucleotides of DNA are melted to expose the template strand; (iii) the template-directed synthesis of an RNA chain begins; and (iv) RNA polymerase clears the promoter and enters a distinct elongation mode [for reviews, see von Hippel et al. (1984) and McClure (1985)].

Successful initiation requires that promoter binding be tight enough to form the open complex, but not too tight to allow the escape of polymerase from the promoter to elongate mRNA. This inherent conflict appears to be expressed in certain irregularities of the initiation process. For example, nascent transcripts up to nine nucleotides long can be released from the RNA polymerase–promoter complex without dissociation of the polymerase from the promoter, whereupon the enzyme starts another chain from the beginning (Johnston & McClure, 1976; Carpousis & Gralla, 1980, 1985; Reznikoff et al., 1982; Spassky, 1986; Levin et al., 1987; Krummel & Chamberlin, 1989). Such "abortive initiation" occurs in vitro at a variety of *E. coli* promoters, and may reflect an activity common to both prokaryotic and eukaryotic RNA polymerase (Yamakawa et al., 1981; Cowie et al., 1982; Luse & Jacob, 1987). For the bacterial enzyme, conversion from a recycling abortive complex to a stable elongation complex has been correlated with loss of the σ subunit after transcription of 9–11 nucleotides (Hansen & McClure, 1980; Straney & Crothers, 1985). To account for abortive cycling and for some heterogeneity in transcription initiation sites, Carpousis et al. (1982) proposed that promoter-bound RNA polymerase chooses its initiating nucleotide and synthesizes short RNA chains from a small initiation region without breaking the original strong contacts that form the open complex.

Here we demonstrate a further variant of the initiation process that produces extensive 5′ heterogeneity of the late gene promoter transcript of bacteriophage 82: the initial trinucleotide pppAAA of the RNA slips against the template TTT, beginning a slippage cycle through which the transcript acquires an untemplated initial sequence of poly(A) that may exceed 20 nucleotides. A similar slippage reaction in a mutant promoter was reported by Harley et al. (1990). This reaction may be similar enzymatically to less natural reactions in which a DNA homopolymer, an RNA homopolymer, or a short homopolymeric DNA sequence within a larger chemically random DNA sequence can serve as template to make a long complementary-sequence RNA homopolymer (Falashi et al., 1963; Stevens, 1964; Chamberlin & Berg, 1964; Fox & Weiss, 1964; Krakow & Karstadt, 1967).

Experimental Procedures

Materials. All materials were of the highest purity available and were purchased from commercial sources, unless otherwise stated. NusA protein was purified by the method of Schmidt and Chamberlin (1984b) with the modification described by Goliger and Roberts (1987). RNA polymerase was purified in this lab by the methods of Burgess and Jendrisak (1975) as modified by Lowe et al. (1979). Radiolabeled nucleoside triphosphates were purchased from Amersham; regular and "ultrapure" (grade 3) nucleoside triphosphates were from

† This work was supported by Grant GM 21941 from the National Institutes of Health.
‡ Present address: Department of Biochemistry and Molecular Biology, 7 Divinity Ave., Cambridge, MA 02138.

0006-2960/90/0429-10702$02.50/0