LBL--31881

DE92 017121

**SMALL CATALYTIC RNA:
STRUCTURE, FUNCTION AND APPLICATION**

Joseph Albert Monforte
Ph.D. Thesis

Department of Chemistry
University of California

and

Laboratory of Chemical Biodynamics
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

April 1991

MASTER

de
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

This work was supported by the Director, Office of Energy Research, Office of General Life Sciences, Structural Biology Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

**Small Catalytic RNA:
Structure, Function and Application**

Copyright © 1991

by

Joseph Albert Monforte

The Government reserves for itself and others acting on its behalf a royalty free, nonexclusive, irrevocable, world-wide license for governmental purposes to publish, distribute, translate, duplicate, exhibit, and perform any such data copyrighted by the contractor.

The U.S. Department of Energy has the right to use this thesis for any purpose whatsoever including the right to reproduce all or any part thereof

Small Catalytic RNA: Structure, Function and Application

by

Joseph Albert Monforte

Abstract

We have utilized a combination of photochemical cross-linking techniques and site-directed mutagenesis to obtain secondary and tertiary structure information for the self-cleaving, self-ligating subsequence of RNA from the negative strand of Satellite Tobacco Ringspot Virus. We have found that the helical regions fold about a hinge to promoting four different possible tertiary interactions, creating a molecule of similar shape to a paperclip. Mutagenesis and other information was used to determine which of the four possible conformations promoted the cleavage and ligation reactions. By using distance geometry techniques we show that such a structure can readily occur without any major steric conflicts. A model suggesting that the "paperclip" and "hammerhead" RNAs share a similar three dimensional structure is proposed.

We have used a self-cleaving RNA molecule related to a subsequence of plant viroids, a "hammerhead", to study the length-dependent folding of RNA produced during transcription by RNA polymerase. Transcript elongation is arrested at defined positions using chain-terminating ribonucleoside triphosphate analogs. When the transcript can form the "hammerhead" structure, it self-cleaves to give a truncated product. The experiment yields an RNA sequencing ladder which terminates at the length at which cleavage becomes possible. We have used this method to determine the length of RNA sequestered within elongating *E. coli* and T7 RNA polymerase complexes. The data show that for *E. coli* RNA polymerase 12 ± 1 nucleotides are sequestered within the ternary complex, which is consistent with the presence of an RNA-DNA hybrid within the transcription bubble, as

proposed by others. The result for T7 RNA polymerase differs from *E. coli* RNA polymerase, with only 10 ± 1 nucleotides sequestered within the ternary complex, setting a new upper limit for the minimum RNA-DNA required for a stable elongating complex. Comparisons between *E. coli* and T7 RNA polymerase are made. The relevance of the results to models for transcription termination, abortive initiation, and initiation to elongation mode transitions are discussed.

A BIM BERI GLASSALA GLANDRID
E GLASSALA TUFFM I ZIMBRA

D. Byrne

Table of Contents

ABSTRACT	1
TABLE OF CONTENTS	iii
ACKNOWLEDGEMENTS	v
GENERAL INTRODUCTION	1
CHAPTER 1. The Structure of the Catalytic RNA from the Negative Strand of Satellite Tobacco Ringspot Virus A "Paperclip" Model	
Introduction	3
Materials and Methods	12
(a) <i>Materials</i>	12
(b) <i>DNA Templates</i>	12
(c) <i>In Vitro Transcription Reactions</i>	13
(d) <i>Catalytic Activity Assay</i>	14
(e) <i>Psoralen Cross-linking</i>	14
(f) <i>Cross-link Purification and Isolation</i>	15
(g) <i>Reverse Transcriptase Sequencing and Cross-link Analysis</i>	16
(h) <i>Computer Modelling</i>	17
Results	18
(a) <i>Cross-link Analysis</i>	19
(b) <i>Reverse Transcriptase Analysis</i>	26
(c) <i>Monoadduct Sites</i>	26
(d) <i>Intermolecular Cross-links</i>	32
(e) <i>Intramolecular Cross-links</i>	36
(f) <i>Site-directed Mutagenesis</i>	40
Discussion	46
<i>"Hammerhead" versus "Paperclip"</i>	55
References	59

CHAPTER 2. Part One: Development of the Transcript Folding Assay and
Application to *E. Coli* RNA Polymerase

Introduction	62
Materials and Methods	67
(a) <i>Materials</i>	67
(b) <i>DNA Templates</i>	67
(c) <i>In Vitro Transcript Reactions</i>	70
(d) <i>Transcript Analysis</i>	71
Results	72
(a) <i>Transcript Self-Cleavage in Stalled Elongation Complexes</i>	74
(b) <i>Temperature Dependence</i>	82
Discussion	86
<i>Relevance to Termination</i>	87
Part Two: Extension of the Transcript Folding Assay to T7 RNA Polymerase	
Introduction	92
Materials and Methods	94
(a) <i>Materials</i>	94
(b) <i>DNA Templates</i>	94
(c) <i>In Vitro Transcription Reactions</i>	94
Results	96
<i>Transcript Self-cleavage in Stalled Elongation Complexes</i>	97
Discussion	101
References	103

Acknowledgements

This work was accomplished with the help of many people. I will do my best to list them all.

First on the list is our fearless leader, John Hearst, who gave me the full freedom to explore the projects that I found most interesting and a wealth of tools to dig with.

Chapter 2 of this thesis represents a true group effort. Without the hard work of Jason Kahn and Kamala “eat more vegetables” Tyagarajan this work would have never manifested so completely.

Thanks goes to John “John John” Hubbard for his patience in teaching me computer modeling while he wrote his thesis, and for trusting me enough to not trash his computer.

My appreciation goes out to David Koh, Ann Caviani Pease, Chris Noren, and unknown members of the Chamberlin lab for providing large amounts of DNA and polymerase. Others I wish to thank include Arnold Hampel, George Bruening, Paul Feldstein, O. Uhlenbeck, John Jaeger, Nacho Tinoco, and members of the Hearst and Tinoco labs, past and present for helpful advice and discussion, and also my lab mate Hans for helping me create an environment pleasing to work in.

I wish to thank my fan club, Frank and Ethel Monforte who probably knew I would be able to do it on the day I was born. They have continually supported me with love, prayers and the occasional check.

Finally and most importantly I must thank my soul mate and love, Tracy Peterson, who makes me feel complete. To her I dedicate this work.

General Introduction

We are now celebrating the tenth anniversary of the discovery of catalytic RNA. In the short history of this new field (it is still called new) a revolution has taken place in thought and understanding of the roles which RNA has in the biological system and in the origins of life. Tremendous excitement exists over the possible uses for a material that can potentially perform many of the same functions as proteins yet is as easily copied and amplified as DNA.

Until Tom Cech discovered catalytic activity within Tetrahymena RNA, it was believed that RNA served only simple well known functions within the cell, involved only in the storage and transport of genetic material or the structural scaffolding of the ribosomes. His molecule defied this classic dogma and performed RNA cleavage and ligation reactions in the absence of protein helpers. Since then the field has expanded to include many classes of catalytic RNA which perform a variety of essential biological tasks. The activities of these catalytic RNAs include, intron excision and exon splicing by Tetrahymena and other Group I introns, tRNA maturation by M1 RNA in RNase P, and the autolytic cleavage and ligation of RNA viral genomes during their replication process. Others have proposed that the RNA within the ribosome plays a large active role in protein synthesis, which suggests that RNA was the biological precursor to the more flexible and adept protein. A tremendous interest exists in answering questions about RNA's role in the origins of life.

In a world dominated by the much more complex functionality of proteins it is easy to think that the potential of RNA is very limited, that four different building blocks versus twenty must lead to dramatically reduced potential. In some ways this point is true, otherwise why would evolution have chosen protein enzymes over RNA enzymes. It is also unlikely that RNA alone will be able to perform some types of reactions, such as

oxidation and reduction. But these limitations do not mean that RNA is incapable of many thousands of different structures and activities. Even a system with one building block, such as Lego™, can be used to construct a near infinite number of different structures with different functionalities.

In order to realize the potential of the RNA world one must start by studying the examples of catalytic RNA nature has provided us. By understanding the structural motifs within these RNAs we will gain a greater understanding about how these molecules activate specific reactions and provide a basis for the design of new RNA based biological tools.

We have focused our studies on two small catalytic RNA structural motifs found within plant viral RNA, these are the “hammerhead” and “paperclip” structures. In chapter 1, several methods are used to obtain structural information from the “paperclip” RNA. These methods include psoralen cross-linking coupled with reverse transcriptase analysis, site directed mutagenesis, and molecular modeling. We use these methods to identify the secondary and tertiary interactions of the self-cleaving “paperclip” RNA and to determine how such a structure promotes site specific cleavage. In addition, we look to see if the “hammerhead” and “paperclip” molecules, which have very little linear sequence homology, share a common catalytic tertiary structure.

In chapter 2, we investigate the use of the self-cleaving “hammerhead” RNA as an assay for RNA structure formation within nascent transcripts during active transcription. Our goal is to study the length-dependent folding of RNA transcripts produced during transcription by various RNA polymerases and to relate it to a physical model of the ternary transcription complex. We also want to address questions of how the complex may vary between different RNA polymerases, and how RNA folding affects different processes such as pausing and termination.

Chapter 1

The Structure of the Catalytic RNA from the Negative Strand of Satellite Tobacco Ringspot Virus A “Paperclip” Model

Introduction:

The ability to self-cleave during the cycle of replication is a common theme among viroids, virusoids, and linear satellite RNA. In early *in vitro* studies of these RNAs it was discovered that a fraction of these small RNA genomes could be isolated as circular forms and linear multimeric dimers, trimers, etc. (Bruening *et al.*, 1982; Kiefer *et al.*, 1982). This discovery suggested that these molecules were likely to replicate using a rolling circle mechanism. A single stranded circular monomer of the genome is utilized by an RNA dependent RNA polymerase to produce a contiguous, complimentary, linear multimer which is then cut up into monomers. Some of these monomers are then circularized and used as a template to transcribe a long, multimeric form of the original genome which is then cut into mature monomers. In the case of viroids and virusoids the monomers are then circularized into their active form. Originally it was thought that this all took place using various polymerases, ligases and nucleases provided by the host cell. As the search for the proteins responsible for replication progressed, it became apparent that not all steps were protein dependent. Prody *et al.* (1986) found that dimeric forms of the (+) strand of Satellite Tobacco Ringspot Virus (STobRV) autolytically cleaved in the absence of any proteins. Soon thereafter this self-cleavage event was found in a whole host of viroids and

solution, undergoes a 3'-endo to 2'-endo conformational switch in the ribose which in conjunction with a rotation of the phosphate backbone produces the catalytically active geometry. Heus and Pardi (1991) have substantiated the 2'-endo conformation of the nucleotide on the 5' side of the cleavage site using two dimensional NMR techniques. The cleavage reaction occurs by transesterification, a 2' hydroxyl attack on the adjacent phosphate displacing the 5' oxygen. The resultant products are a 2',3'-cyclic phosphate and a free 5' hydroxyl (Prody *et al.*, 1986). The reaction proceeds at neutral pH and requires a strong cationic environment. Activity is optimal in the presence of magnesium ion though it will occur at high sodium ion concentrations. The reaction requires no cofactor or external energy source.

The "hammerhead" is not the only self-cleaving RNA motif to rise out of the viral RNA kingdom. A completely different self-cleaving RNA structure occurs in (-) Satellite Tobacco Ringspot Virus. This sequence which shares no homology with the "hammerhead" motif was found to not only autolytically cleave but to also ligate at one specific site (Buzayan *et al.*, 1986). On a microscopic level the mechanism for cleavage is identical, producing the same 2',3'-cyclic phosphate and free 5' hydroxyl end products. Because the transesterification is energy neutral it is not surprising that the products can also religate. Entropy factors strongly in the ligation reaction and for the most part yields are low. In theory, the "hammerhead" geometry should also religate, and as a small unimolecular system it has been shown that it does (George Bruening, personal communication). As with the "hammerhead" sequence, truncation experiments have been performed to reduce (-)STobRV to the shortest sequence capable of cleavage (Feldstein *et al.*, 1989; Hampel and Tritz, 1989; Haseloff and Gerlach, 1989). The result is a molecule with at least 56 nucleotides in size which contains sequences from two discontinuous sections of (-)STobRV. Figure 2 shows a computer generated secondary structure for the minimum autolytic sequence. This larger "minimum" sequence shows the highest level of cleavage activity. The nucleotides which are boxed may also be eliminated but the level of

activity is noticeably reduced. This abbreviated bimolecular system religates with very low efficiency (George Bruening, personal communication).

Figure 2. Self-cleaving Structure from (-)STobRV

Computer generated secondary structure truncated subsequence from the negative strand of Satellite Tobacco Ringspot Virus with the maximum rate of cleavage activity. Structure without the blocked sequence is the smallest sequence capable of self-cleavage. Two possible pairings for helix 4 are shown.

This structure appears to be very different from the "hammerhead". There is no three helix junction, instead the structure is composed of a set of four helices. The two inner helices, 2 and 3, are joined on only one side providing additional degrees of freedom. The end helices, 1 and 4 are separated from the inner helices by internal loops which have undefined structure. Of the four helices, helix 4 is the least defined. Two different foldings are produced by the secondary structure folding program (Fig. 2) both of which have only a few base pairs of uninterrupted helix, suggesting it may have only minimal stability.

Two new sequences which share a high level of homology with (-)STobRV have been found, one in (-) Satellite Arabis Mosaic Virus (Kaper *et al.*, 1988), and the other from (-) Satellite 1 of Chicory Yellow Mottle Virus (Rubino *et al.*, 1990), shown in Figure 3.

Figure 3. Self-cleaving Sequences Homologous to (-)STobRV

Subsequences from the negative strands of Satellite Arabis Mosaic Virus and Satellite 1 of Chicory Yellow Mottle Virus. Nucleotides which differ from Satellite Tobacco Ringspot Virus are highlighted in bold type.

Sequence variations from (-)STobRV are shown in bold. It is interesting to note that the internal loop sequences have been nearly completely conserved. The exceptions all occur at one end of each of the two internal loops. Preliminary evidence has shown that both sequences as part of the full size molecules, 300 and 457 nucleotides respectively, undergo self-cleavage. Neither sequence has yet been reduced to the minimum size shown here, though both are expected to retain cleavage activity in the truncated form.

Hampel *et al.* (Hampel *et al.*, 1990) have performed extensive mutagenesis on the truncated (-)STobRV. Mutant revertant experiments have firmly established the presence of all of the base pairs in helices 1 and 2, and two of the base pairs in helices 3, and 4. In addition, the adenosine on the 5' side of the cleavage site was shown to be completely variable. Mutations to the three nucleotides on the 3' side of the cleavage site have resulted in a loss of activity. A summary of their results are shown in Figure 4A. Insertion deletion studies by Haseloff and Gerlach (1989) demonstrated that the minimum (-)STobRV structure could be interrupted with two different large insertions yet still maintain activity,

these insertions are shown in Figure 4B. These are very large perturbations in a relatively small system and any complete model for this molecule must explain their ineffectiveness.

Figure 4. Mutagenesis Consensus and Insertions Into (-)STobRV

A: Consensus of mutagenesis data. N represents variable nucleotide sites. B: Two different large insertions which do not destroy cleavage activity.

Both the “hammerhead” and the alternate structure from (-)STobRV can be created as uni- or bimolecular systems. As bimolecular systems they both can act as RNA enzymes, ribozymes, to catalyze cleavage of an RNA substrate. Both follow Michealis-Menten kinetics rules and have an optimal rate of ~2 per minute (Hampel and Tritz, 1989; Uhlenbeck, 1987). The main difference is that the former runs optimally at 55 °C while the latter at 37 °C, which is attributed mostly to the kinetics of helix hybridization rather than some mechanistic detail of the transesterification reaction.

The intentions of the experiments and analysis presented here are to provide a more complete picture of the (-)STobRV ribozyme and how it interacts with its substrate to catalyze cleavage. Specifically, we wish to identify the interactions formed by the internal loop sequences and how they might participate in catalytic function. It is not readily apparent what interactions may be occurring within these regions. Secondary structure

predictions show a tremendous variability in the kinds of base pairings that might exist, and cannot predict what tertiary interactions these regions might be involved in. There is also a very real possibility that the nucleotides in these regions may be involved in unconventional bonding interactions, such as nonstandard base pairing, base-sugar or base-phosphate interactions. Parts of these regions may be involved in no interactions at all. In addition, we want to further establish the presence and importance of different elements within helix 4. With this knowledge in hand we intend to create a three dimensional model showing how this molecule accomplishes its job.

The methods used are three-fold: psoralen cross-linking to establish the presence of tertiary interactions, site-directed mutagenesis to assay the importance of different interactions to catalytic function at nucleotide resolution, and computer modelling as a tool to assess the feasibility of the molecule to form different structures without excessive steric conflicts, and provide a visual model to aid our intuitive process.

The specificity of the psoralen photoreaction with double stranded nucleic acid has led to their widespread use as human therapeutics and as probes of nucleic acid structure. The reactivity of psoralens with nucleic acids has been well studied (Cimino *et al.*, 1985). Psoralens are able to intercalate into double stranded regions of nucleic acid. Upon irradiation with long-wave ultraviolet light the intercalated psoralens are able to photoreact with adjacently stacked pyrimidine bases. This reaction occurs both in DNA and RNA. Within RNA, uridines are more reactive than cytosines by a factor of 10. Photo-cross-linking is a two step process, by absorbing a single photon psoralen can react with an adjacent pyrimidine to give a monoadduct. A cross-link is formed when the psoralen monoadduct absorbs a second photon and reacts with a second pyrimidine, opposite and adjacent to monoadducted pyrimidine. Both the monoadduct and cross-link are photoreversible upon irradiation with UV light of the appropriate wavelength.

Psoralen's ability to cross-link nucleic acid makes it an excellent probe for RNA secondary and tertiary structure, and it has been used extensively to identify various

structural elements in molecules such as ribosomal RNA, M1 RNA and small nuclear RNA (Ericson and Wollenzien, 1989; Lipson *et al.*, 1988; Maser and Calvet, 1989). Key to the psoralen cross-linking method are the techniques used to identify the sites of adduction. The three main methods that have been used, in order of increasing resolution, are, electron microscopy, partial RNase digestion, and reverse transcriptase analysis. Because the molecule of interest is relatively small we have opted for the reverse transcriptase method which has the highest resolution of the three methods, and can identify adduction sites to nucleotide resolution. Reverse transcriptase, which makes cDNA from an RNA template, stops at psoralen lesions in the RNA template. Youvan and Hearst (1982) have identified psoralen photoreactive sites in 16S rRNA using reverse transcriptase to make abbreviated cDNA transcripts corresponding to termination at psoralen adducts (both monoadducts and cross-links). Ericson and Wollenzien (1988) further refined and characterized the method. Photoreacted RNA molecules were first purified by denaturing gel electrophoresis, separating the different cross-link species from each other and from monoadducted and unreacted RNA. These purified cross-links are then subject to reverse transcriptase analysis, resulting in only a few stops and allowing for the unambiguous assignment of different cross-link locations. They found that reverse transcriptase stops primarily one nucleotide in front of the site of adduction and secondarily directly on the site of adduction. Psoralen lesions, both monoadduct and cross-link, are absolute stops for reverse transcriptase, which cannot proceed even after long periods of incubation.

Site-directed mutagenesis was used to answer a variety of questions. Most mutations made were point mutations or deletions targeted to the internal loop regions aimed at determining the importance of each nucleotide in the overall catalytic structure and the type of interaction it is involved in. Several mutant revertant pairs were made to assess the existence of specific base pairs. Other mutants include extensions of helix 4 in various configurations, to establish the importance of different features of that helix. Interpretation of the mutagenesis results was complicated by the multitude of interactions nucleotides can

make within and outside of internal loop structures and how a change in one nucleotide can alter the whole system in unforeseen ways.

The principle method for translating the physical data into a three dimensional model was distance geometry. This program produces structures using a set of atomic distance constraints determined experimentally or otherwise for a given molecule. The structures which can conform to these distance constraints are solved by creating an N dimensional distance bounds matrix, containing upper and lower bounds for all known distances, and then solving it for three dimensional space. The precision of the structure is determined by the number and precision of the initial constraints. This method has been used to predict the global conformation and general backbone path for transfer RNA and 16S RNA (Hubbard and Hearst, 1991).

Our results show that two halves of the (-)STobRV structure bend about a hinge between helices 2 and 3 allowing for a variety of interactions to occur between the two internal loop sections. Modelling shows that these tertiary interactions do not cause excessive van der Waals conflicts, and provides a picture of the molecule that looks not unlike a twisted paperclip.

Materials and Methods:

(a) Materials

Purified T7 RNA polymerase was generously provided by Profs. David Wemmer and Peter Schultz and the hard labor of Ann Caviani Pease and Chris Noren (University of California, Berkeley). Ribonucleoside triphosphates (NTP's), deoxyribonucleoside triphosphates (dNTP's), dideoxyribonucleoside triphosphates (ddNTP's), and ribonuclease inhibitor (RNasin) were purchased from Pharmacia. [γ - ^{32}P]Adenosine 5'-triphosphate and [α - ^{32}P]Uridine 5'-triphosphate were from Amersham or New England Nuclear. Cloned T4 polynucleotide kinase was supplied by New England Biolabs, and AMV Reverse Transcriptase by Life Sciences, Inc. 4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT) and [^3H]HMT were generous gifts from HRI Inc. Electrophoresis grade 40% 19:1 acrylamide:bis(acrylamide) solution was purchased from IBI Inc. Ammonium persulfate, and TEMED were obtained from Bio-Rad. Ultra pure urea was from Schwarz/Mann Biotech. Diethylpyrocarbonate-treated water, silanized Eppendorf tubes, and deionized formamide were used throughout.

Radioactivity assays were performed using an LKB 1209 Rackbeta Liquid Scintillation Counter by either Cherenkov counting or in the presence of 5 ml of ICN Ecolume fluor. All gel autoradiography exposures were taken with Kodak XAR-5 X-ray film and Cronex Lightning Plus enhancing screens (DuPont).

(b) DNA Templates

Deoxyoligonucleotides (15–96 nt) were synthesized on an Applied Biosystems 381A automated synthesizer and purified by gel electrophoresis on 10–20% polyacrylamide gels. T7 RNA polymerase templates were constructed by hybridizing synthetic oligonucleotides, providing an 18 nucleotide long double stranded promoter sequence followed by a long, single stranded template sequence (Milligan *et al.*, 1987). The templates have an abbreviated, strong Class III promoter immediately followed by the template transcription

sequence, coding for either the ribozyme or ribozyme substrate as shown in Figure 5. Mutant sequences vary only by the specific point changes, additions or deletions.

Ribozyme Template

```

5' CTAATACGACTCACTATA 3'
3' GATTATGCTGAGTGATATCCTTTGTCTCTTCTGTTGGTCTCTTTGTGTGCAACA...
      T7 Promoter      |+1

...CCATATAATGGACCATGCGGAAGATC 5'

```

Substrate Template

```

5' CTAATACGACTCACTATA 3'
3' GATTATGCTGAGTGATATCCACAGTCAGGACAAAGCAGGAGTGCCTGAGTG 5'
      T7 Promoter      |+1

```

Figure 5. DNA Templates for Synthesis of Ribozyme and Substrate

(c) In Vitro Transcription Reactions

All transcription reactions were performed at 37 °C in a buffer containing 40 mM TrisOAc, pH 8.0, 28 mM MgCl₂, 4 mM NaCl, 2 mM spermidine, 30 mM DTT, and 0.01% Triton-X. Ribonucleoside triphosphate concentrations were all 4 mM, and DNA template 1 μM. [α -³²P]UTP (3000 Ci/mmol) was used to body label and was incorporated at a level of ~1:20,000 [α -³²P]UTP:UTP. Reactions were initiated by the addition of 5 U/μl of T7 RNA polymerase and 1 U/μl of RNasin, then incubated for 2 hours. Transcription was terminated by the addition of an equal volume of a formamide/loading dye solution and then loaded directly onto a 15% denaturing polyacrylamide gel. Body-labeled RNA was detected by autoradiography with typical exposure times of 5–10 minutes. RNA was then cut from the gel and eluted into 4 ml of TE buffer solution, (10 mM TrisOAc, pH 7.5, 1 mM EDTA), and precipitated by the addition of 10 mM MgCl₂ and 2 volumes of cold EtOH. Specific activity was used to quantitate the RNA and product identity was verified by reverse transcriptase sequencing and/or an assay for catalytic activity, both of which are described below.

(d) Catalytic Activity Assay

Catalytic activity was assayed by incubating a mixture of ribozyme and substrate in a buffer composed of 40 mM TrisOAc, pH 8.0, 12 mM MgCl₂, and 2 mM spermidine at 37 °C. Ribozyme and substrate were added at concentrations that ranged between 10-500 nM, but always kept at a stoichiometric ratio of 1.2:1, ribozyme:substrate. Time course incubation times ranged from 30 seconds to 4 hours, and were quenched by the addition of a mixture of formamide, 15 mM EDTA, and loading dye. Samples were concentrated in a Speed-Vac (Savant) to 5 µl, and loaded directly onto a 10% denaturing polyacrylamide gel. Gels were exposed to X-ray film to locate RNA samples which were excised and quantitated by liquid scintillation. Percent cleavage of substrate was calculated by taking the ration of cleaved to uncleaved substrate at each time point and adjusting for background at the zero time point.

(e) Psoralen Cross-linking

100 pmol of ribozyme was mixed with saturating HMT, 36 µg/ml, in the catalytic activity buffer listed above. Prior to irradiation 100 pmol of substrate was added to the mixture bringing the volume to 100 µl. For samples of ribozyme alone, only water was added. After four minutes of incubating at room temperature, allowing for complete mixing of RNA with psoralen, the reaction mixture was then irradiated with UV light. Following irradiation the cleavage activity was quenched by the addition of excess EDTA. Additional experiments allowed the RNA to incubate for an extended period of time prior to or following irradiation but before quenching. Typical experiments used between 800-1600 pmol of RNA per reaction condition.

Two different light sources were used for psoralen photochemistry. The first was a Mercury/Xenon arc lamp filtered through water and cobaltous nitrate solutions to produce 300 mW/cm² of 320 to 380 nm UV light. Samples were irradiated for 5 minutes. The second light source was a high intensity 366 nm Argon laser (Spectra Physics), running at

2-3 W intensity, allowing for shorter, 30 second irradiations. In both cases, samples were exposed to enough UV light to completely consume all of the photoreactive psoralen present in solution. Psoralen adduction levels were held to one adduct per 2 RNA complexes in order to limit complications in the reverse transcriptase analysis which will be discussed later. Because the majority of psoralen adducts are monoadducts and not cross-links the yield of cross-linked RNA is very low, between 0.1 and 1.0%.

Cross-linked RNA reaction mixtures were dialysed and concentrated using Centricon 10 (Amicon) microdialysis tubes. Samples were spun at 5000g in a Sorval SS-34 rotor for 90 minute intervals. Two 2 ml washes with TE were performed to reduce salt concentrations to a minimal level. Sample volumes were reduced to between 50 and 100 μ l and combined with an equal volume of a formamide/loading dye. Microdialysis was used instead of precipitation because of problematic enhanced aggregation that occurs with precipitated, psoralen adducted RNA.

In order to quantitate the average amount of psoralen incorporation, photoreactions were done using [3 H]HMT in place of cold HMT under identical conditions to those above. Following irradiation, excess [3 H]HMT was dialysed away in three 2 ml washes of water using Centricon 10 microdialysis tubes as above. The amount of HMT and RNA were then quantified by 3 H scintillation and UV absorbance respectively.

(f) Cross-link Purification and Isolation

The principle method for cross-link separation was denaturing gel electrophoresis. Samples were boiled for 1.5 minutes, cooled on ice, and loaded directly onto 0.5 or 1.2 mm thick, 33 $\times$ 43 cm, polyacrylamide (1:19 bis:acrylamide) gels containing 8 M urea and 0.6 $\times$ TBE (50 mM Tris, 50 mM borate, 1 mM EDTA, pH 8.3). Gel percentages varied between 12-20% and were run at two different power levels, 55 W, ~2300 V, ~50 $^{\circ}$ C or <1 W, ~500 V, 37 $^{\circ}$ C. High temperature and low wattage yielded the best resolution. Cross-links were detected by autoradiography with typical exposure

times of 12–24 hours. Isolated cross-links were cut from the gels, eluted into 8 ml TE and then concentrated using Centricon 10 microdialysis tubes. Several inter- and intramolecular cross-links which comigrated in the first round of electrophoresis were rerun on a second 20% gel at 4 W, –1000 V, and then purified in an identical manner as the first round.

(g) Reverse Transcriptase Sequencing and Cross-link Analysis

Two 15 nt long DNA primers complementary to the 3' ends of the ribozyme and substrate were kinased in a standard buffer (Maniatis *et al.*, 1982) for one hour at 37 °C using [γ -³²P]ATP. Typical reaction mixtures included 250 μ Ci label (5000 Ci/mmol), 240 pmol DNA primer, and 40 U kinase in 60 μ l. The kinase reaction was heat-killed for 3 minutes at 70 °C.

For each reaction 4 pmol of the kinased primer were annealed to 400 fmol of RNA. Sample volume was reduced to a 3 μ l buffer containing 250 mM KCl and 10 mM TrisOAc, pH 8.3, heated to 100 °C for 2 minutes, and then allowed to slowly cool to room temperature.

Reverse transcription was initiated by combining the annealed mixture with a 3.3 μ l solution containing 24 mM TrisOAc, pH 8.3, 16 mM MgCl₂, 100 mM KCl, 2 mM spermidine, 16 mM DTT, 500 μ M dATP, dCTP, dTTP, 800 μ M dGTP, 6 U RNasin, and 20 U AMV reverse transcriptase. For sequencing reactions the reverse transcription mix also included one of four ddNTPs at a 300 μ M concentration. Reactions were incubated at room temperature for 10 minutes and then at 50 °C for 50 minutes. Reactions were quenched by the addition of 4 μ l of formamide/loading dye solution.

Samples were concentrated in a Speed-Vac (Savant) to 4 μ l, boiled for 1.5 min, cooled on ice, and loaded directly onto 0.4 mm thick, denaturing, 10% polyacrylamide gels. Gels were run until Bromophenyl Blue reached the bottom. Typical autoradiography exposure times were 20 minutes to 4 hours.

(h) Computer Modelling

Secondary structure predictions were performed using MFOLD, ver. 2.0, which uses free energy minimization to determine optimal and suboptimal secondary structure foldings for RNA at different temperatures (Jaeger *et al.*, 1989a; Jaeger *et al.*, 1989b; Zuker, 1989).

Distance geometry was used to determine the global shape and backbone path for the different possible RNA conformations. The program, DSPACE, ver 2.1, was written by Dennis Hare and Robert Morrison at Hare Research, Inc. Distance constraints for helical regions and for the base paired tertiary interactions were approximated using coordinates from fiber diffraction data of RNA A-form helix (Arnott *et al.*, 1973). Reduced atom representations were used to represent the helical and single stranded regions because the RNA molecules studied contain too many atoms, over 2300, to be modeled within a reasonable amount of time and using an affordable amount of memory. The types of pseudoatom and pseudohelical representations used are detailed in Hubbard and Hearst (1991) as are the justifications for choosing these particular representations. All computations were performed on a MicroVAX II. The final structures were expanded to full representation using modules from AMBER, ver. 3.0 (obtained from Peter Kollman at UCSF), visualized on an Evans & Sutherland PS340, and then converted to black and white traces on a Macintosh II. Line drawings of the reduced atom molecules produced by DSPACE were visualized on terminals emulating Tektronix video displays.

Results:

The goal of this work was to obtain a more detailed picture of the catalytically active subsequence of (-)STobRV RNA. While previous work has determined the primary sequence and most of the secondary helical domains (Hampel *et al.*, 1990), nothing was known about what tertiary interactions might be involved in the catalytic structure. Although the catalytic portion of the molecule is relatively small, approximately 70 nucleotides for a minimal active structure, it is still too large to be easily deciphered by NMR and therefore required the use of more indirect biochemical methods. The methods we chose to approach this problem were psoralen cross-linking and site-directed mutagenesis. Psoralen cross-linking was used to provide a hard physical map of tertiary interactions which are not easily identified by other methods, in addition to some secondary interactions. Site-directed mutagenesis was used to fine tune the structure by identifying the importance of contributions by individual nucleotides and small sequential motifs to the overall structure, and to differentiate between active and inactive conformations.

Prior to the synthesis of any physical RNA the MFOLD secondary structure folding program was used to provide an educated starting point about the potential structure(s) that the molecules could form, producing the picture in Figure 6A. The internal loop regions have been left single stranded, not because they are likely to be single stranded but because their base pairings are not well determined. The secondary structure folds show a variety of pairings for these regions, all with very similar free energies.

All experiments were performed using a bimolecular system, the RNA was comprised of a ribozyme and a substrate modelled after the sequences used by Hampel *et al.* (1989) (Fig. 6B). RNA polymerase reactions and RNA self-cleavage have similar magnesium requirements, with a unimolecular self-cleaver one reaction cannot be performed in exclusion of the other. The bimolecular system allows complete control over the start of the cleavage reaction which was essential for rate studies and for large scale synthesis. The 3' ends of the ribozyme and substrate were extended 15 nt beyond the lengths found in the

Figure 6. (-)STobRV Minimum Sequence and Sequence Used for Cross-linking

minimum active structure. These extensions, which are continuations of the native (-)STobRV RNA sequence, are required to allow the hybridization of DNA primers for post cross-linking reverse transcriptase analysis. A single base pair of helix 2 was switched in order to perform other experiments which are not described here. These changes had no significant effect on catalytic activity.

(a) Cross-link Analysis

All photochemistry was performed in the presence of 12 mM MgCl₂, the optimum concentration for RNA cleavage (Hampel and Tritz, 1989), in order to enhance the catalytically active population. The RNA psoralen mixture was given four minutes of incubation time allowing for complete mixing to occur and to ensure that the cleavage reaction was well on the way. Three types of mixtures were cross-linked: ribozyme alone, to determine intramolecular cross-links, ribozyme with substrate, and ribozyme with substrate where the ribozyme contains no radioactive label, visualizing only intermolecular cross-links. The substrate did not form any major cross-links with itself.

Figure 7 shows the first round of purification of the cross-linked RNA. The left lane is radioactively labeled ribozyme in the absence of substrate. The center lane is a mix of labeled ribozyme and substrate, and the right lane is the reaction of unlabeled ribozyme with labeled substrate. Overall there are 15 distinct cross-link bands migrating above the ribozyme. What is immediately obvious is that the upper bands, in the center and right lanes, are intermolecular cross-links between ribozyme and substrate and the lower bands, in the left and center lanes, are intramolecular cross-links within the ribozyme itself. In the intermediate region, bands 5-7, there is a mixture of inter and intramolecular cross-links which are co-migrating together.

A second, nearly identical gel (not shown) loaded with aliquots of same three cross-link reactions was run for a shorter period of allowing for the collection of unreacted and mono-adducted substrate, which ran off the end of the gel in Figure 7, in addition to the cross-links. The overall cross-link pattern was the same as the first gel in Figure 7, but individual cross-links were not as well resolved from each other. These cross-links were therefore analyzed separately (lanes in Figures 9 and 10 labeled with a †).

In order to separate the mixed cross-links, bands 5-7, a second, higher percentage, denaturing gel was run, 20% versus 12%. In addition to bands 5-7, bands 1-4 were rerun because a small but significant amount of unreacted and mono-adducted ribozyme smears up into the cross-link region. Figure 8 shows the results of the second cross-link purification gel. What is seen is that the banding pattern of the ribozyme alone lanes, is repeated in the ribozyme plus substrate lanes, with the addition of some faster migrating bands in rib+sub lanes 5-7.5. These faster migrating bands are the intermolecular cross-links and migrate at the same distance as the unlabeled ribozyme plus labeled substrate cross-links in the UR+S lanes. It is unusual for these intermolecular cross-links, which have a higher molecular weight, to migrate faster than the smaller intramolecular cross-links. As the reverse transcriptase analysis will show these cross-links all occur with cleaved substrate, which significantly alters their hydrodynamic properties. Another

Figure 7. *Cross-link Purification: First Round*

Purification of three different photoreactions. Ribozyme alone: [α - ^{32}P]labeled ribozyme was cross-linked in the absence of substrate producing intramolecular cross-links isolated in bands 1 through 8r. Ribozyme + Substrate: A 1.2:1 ratio of [α - ^{32}P]labeled ribozyme and substrate were photoreacted to produce a mixture of intra and intermolecular cross-links isolated in bands 1 through 15. Unlabeled Ribozyme + Substrate: Same as R + S except that cold ribozyme was present in the reaction mix, allowing for the visualization of intermolecular cross-links isolated in bands 5 through 15. A mix of unreacted and monoadducted ribozyme was isolated in bands R and RM.

Figure 7

Figure 8. *RNA Cross-link Purification: Second Round*

Cross-linked RNA isolated as bands 1 through 7 in the first round of purification were rerun on a higher percentage, denaturing gel. All of the bands in the Ribozyme alone lanes are intramolecular cross-links. All of the bands in the UR+S (Unlabeled Ribozyme + Substrate) lanes are intermolecular cross-links, UR1 through UR3. The Ribozyme + Substrate lanes contain both intramolecular cross-links, bands 1 through 7, and intermolecular cross-links, bands X1 through X4. Contaminating unreacted and monoadducted ribozyme was separated from the cross-links, band R.

RNA Cross-Link Purification: Second Round

Figure 8

interesting observation is that many of the bands which all migrated very differently on the first gel, migrate very closely on this second higher percentage gel, indicating they are likely to be cross-links of very similar structures. The reverse transcriptase analysis will demonstrate this observation to be true.

Under ideal conditions a purified RNA cross-link should have only one psoralen adducted to it. Achievement of such a goal would require a very low level of psoralen adduction and use more RNA than is practical to produce at this point in time. In order to yield enough of each cross-link to allow for reverse transcriptase analysis, 0.8 pmol, given the amount of RNA starting material, 1600 pmol, the level of adduction had to be raised to 50%, or 1 adduct per 2 RNA complexes. The main reason for such a large level of adduction is that cross-link formation is very inefficient. While the psoralen readily reacts with RNA to form monoadduct, only about 1% or less of the monoadduct is then driven on to cross-link. The downside of needing to be at such a high level of adduction, in order to obtain enough of each cross-link, is that there is also a significant amount of double adduct that occurs (RNA complexes containing 2 psoralens). Because the cross-link efficiency is so low the overwhelming majority of the double adducted complexes will be double monoadducts. Similarly for the complexes containing a cross-link, those that are double adducted will have a cross-link and a monoadduct.

If we consider the addition of each psoralen to be a stochastic process which distributes according to Poisson rules, at a 50% level of adduction, we would expect 30% of the RNA to have one adduct and 7.6% of the RNA to have two adducts. If we ignore all the molecules that contain more than two adducts, all with levels less than 2%, the total amount of material that might contain a cross-link of interest is 37.6%. These numbers imply that about 20% ($7.6\% \div 37.6\%$) of the cross-linked RNA will also be monoadducted. This number is far from ideal, but the overall significance of these monoadducts is attenuated by the fact that for a given purified cross-link the monoadducts are distributed over a large number of reactive pyrimidine sites, 10 or more, on the molecule. In other words, for any

given cross-link a monoadduct can be expected to occur at any particular pyrimidine site one tenth of 20% or 2% of the time, providing a 50 fold discrimination between cross-link sites and monoadducts.

(b) Reverse Transcriptase Analysis

Reverse transcriptase makes cDNA copies of the RNA molecules. When the polymerase reaches the point where the RNA has been modified it stops, resulting in the build up of cDNA measuring to the site of modification. Reverse transcriptase stops one nucleotide upstream from the site of a psoralen adduct, though there are instances where the polymerase will advance to the adducted nucleotide. The stop is absolute.

A ^{32}P labeled DNA primer is hybridized to the 3' end of either the ribozyme or the substrate. This primers is extended using reverse transcriptase to the site of modification. The stops are mapped by comparing their cDNA length with standard chain termination sequencing. Because the enzyme cannot bypass the adduct, any modifications downstream from the first adduct cannot be seen. Therefore, in the case of intramolecular cross-links this method only allows the 3' end of the cross-link to be identified.

(c) Monoadduction sites

In theory, any pyrimidine which psoralen can intercalate next to is a site for monoadduction. In actuality, individual electronic and geometric constraints result in a wide variation in reactivity. For example, it has been found experimentally that uridine residues react with psoralen 10 times better than cytosine residues (Cimino *et al.*, 1985). Sterics and geometry are especially significant. Typical "hot spots" for psoralen monoaddition are at helical ends or internal loops which provide a more flexible environment, while maintaining base stacking, allowing psoralen to more easily search conformational space for the correct reactive geometry. The overall result is that some sites become "hot spots" for photoreaction, while others remain completely unreactive.

Following cross-link separation, the uncross-linked ribozyme and substrate were collected as several fractions. Because monoadducts do not significantly alter the geometry of larger RNA molecules they are not separated from native RNA by gel electrophoresis. Therefore, the purified uncross-linked fractions contain a mix of unreacted and monoadducted RNA. Identification of various monoadducted sites was performed using the reverse transcriptase assay.

Figure 9, lanes R and RM, show the sites for monoaddition on the ribozyme. In addition to stops which also occur in the control, Con, lane there are several stops indicating "hot spots" for monoadduction. The stops occur at U33, G35 and A46. The U33 stop is the result of reverse transcriptase stopping on the monoadduct at U33 and the stop at G35 is the result of reverse transcriptase stopping one nucleotide before U34. Presumably reverse transcriptase also stops at U34 because of the monoadduct at U33, but it is not visible because this location is also a pause site for the polymerase. The stop at A46 is due to the presence of a monoadduct at U44. Monoadducts may also be present at C6, C15, U39 and U41, but the presence of similar stops in the control lane prevent confirmation of this.

No clear monoadducts are visible within the substrate (Fig. 10). There is a stop that occurs in the cleaved substrate lane, SC, at U9. However, this may be an artifact of premature termination at the end of an RNA molecule.

With the exception of the stop at U33, all other stops by reverse transcriptase described in this chapter occur one nucleotide before the site of adduction. Henceforth, all stops will be described by the nucleotide that the adduction is attributed to rather than the unmodified nucleotide where the enzyme stops.

Figure 9. *Intermolecular Ribozyme Cross-link Analysis*

Reverse transcriptase analysis of cross-link sites on the ribozyme. R and RM lanes are analysis of unreacted and monoadducted ribozyme. Numbered lanes are analysis of bands from the first round of cross-link purification († lanes are cross-links from a second, unshown first round purification gel), see Figure 7. Lanes X1 through X4, and UR1 through UR3 analyze intermolecular cross-links isolated during the second round purification, see Figure 8. Sequencing lanes are presented as their representative complementary RNA position rather than the cDNA sequence itself. Major cross-link stops are indicated by the labelling on the right side of the figure.

Intermolecular Ribozyme Cross-link Analysis

Figure 9

Figure 10. *Intermolecular Substrate Cross-link Analysis*

Reverse transcriptase analysis of cross-link sites on the substrate. S, SC and S+ lanes are analysis of unreacted and monoadducted substrate. Numbered lanes are analysis of bands from the first round of cross-link purification († lanes are cross-links from a second unshown first round purification gel), see Figure 7. Lanes X1 through X4, and UR1 through UR3 analyze intermolecular cross-links isolated during the second round purification, see Figure 8. Sequencing lanes are presented as their representative complementary RNA position rather than the cDNA sequence itself. Major cross-link stops are indicated by the labelling on the right side of the figure.

Intermolecular Substrate Cross-link Analysis

Figure 10

(d) Intermolecular Cross-links

The analysis of the intermolecular cross-links are shown in Figures 9 and 10. In the best of all worlds each different cross-link would be purified from every other. The resulting reverse transcriptase analysis would show a single stop for where the cross-link was on the ribozyme and a single stop for where it was on the substrate. What one actually sees are many stops, indicative of the copurification of families of similar cross-links, or in some cases, unrelated cross-linked molecules that happen to share similar hydrodynamic properties. An example of this phenomenon is the comigration of inter and intramolecular cross-links in the first purification gel.

The analysis is further complicated by the continued contamination of unreacted, monoadducted, and double adducted RNA. The presence of unreacted and monoadducted RNA is indicated by the presence of full length product in some of the cross-link lanes. The problem is most prevalent in the ribozyme analysis. Substrate and substrate cross-linked to ribozyme migrate very differently on the purification gels reducing the chance of contamination. The contamination problem persists even a second round of purification, as is the case for ribozyme lanes X2 and X3.

The best identified intermolecular cross-links are those in lanes 10 and 13 of the Figures 9 and 10. Lane 10 shows substantial stops at U41 and U44 of the ribozyme. A significant stop also occurs at U39. The corresponding stop on the substrate is at U9. Lane 13 differs by having only one major stop on the ribozyme at U44, and a minor stop at U41. There are now two stops on the substrate, at U9 and C6. This cross-link data has been replicated in several experiments and is of the highest confidence. A summary of these primary cross-links is shown in Figure 11 and 12. There are several points to note in the details at the bottom of Figure 11. The first is that conformation B would suggest that U43 on the ribozyme should be cross-linkable to U9 on the substrate, yet no cross-link

Four different tertiary conformations established by cross-linking

Figure 11. Intermolecular Cross-link Analysis – Lane 10

Nucleotides highlighted in bold type are psoralen reactive sites. Cross-links are represented by the ~ symbol.

Figure 12. Intermolecular Cross-link Analysis – Lane 13

Nucleotides highlighted in bold type are psoralen reactive sites. Cross-links are represented by the ~ symbol.

has ever been observed to this site. The second item is that the structures show a C-U base pair. It is proposed to exist because the cross-link requires the two covalently bonded pyrimidines to be stacked opposite and adjacent from each other in a standard helical configuration. The only way to place U44 adjacent to U9 and still maintain any standard base pairing is to have a C-U base pair be present. The conformations shown in C and D suggests that stem 4 is not completely stable as one might guess for such a short stem.

The X and UR series of lanes, the clearest of which are X1 and UR3, exhibit similar banding patterns to lanes 10 and 13. The key difference is a large stop at G8 on the substrate. This stop is indicative of run-off at the cleavage site, and the attributed cross-links occur with cleaved substrate. Because we see the cleavage run-off product, it means that not all the cross-links are to U9 of the substrate. If that were the case all of the reverse transcriptase would stop at U9 and no run-off product would be seen at all. By analogy to lanes 10 and 13, the likely candidate for the other substrate cross-link site is C6. The fact that cleavage has occurred in these molecules is not conclusive evidence that one of these cross-linked structures is the active conformation, since the cleavage could have occurred before the cross-link was formed.

While the other cross-link lanes are not as clear, they do provide some additional information. In example, ribozyme lane 9 has strong stops occurring at U43, U36, U34, and U33, in addition to the stops at U39, U41, and U44. The substrate lane 9 has stops at U3 and U5, along with U6. Among these stops are cross-links. Besides the previous mentioned cross-links another possibility is a cross-link between U33 or U34 of the ribozyme and U9 of the substrate, though mutagenesis data presented later discounts this interaction as part of the catalytically active structure. At the very least, all the sites listed above are sites of monoadduction. Figure 13 sums up these observations, psoralen reactive sites are bold and regions involved in possible tertiary interactions are indicated.

Figure 13. *Other Possible Intermolecular Cross-links & Monoadduct Sites*

Nucleotides highlighted in bold type are psoralen reactive sites. Cross-links are represented by the ~ symbol.

(e) Intramolecular Cross-links

Due to the shortcomings of the reverse transcriptase analysis complete identification of intramolecular cross-link sites cannot be made directly. What can be identified directly are the downstream ends of each cross-link. The upstream ends must be deduced by taking into account previously known data and the rules for cross-linkable sites. Some cross-link possibilities are suggested below.

There are several potential intramolecular structural motifs that are suggested by the reverse transcriptase patterns (Fig. 14). Similar to the intermolecular data there are strong stops at U39, U41, U43, and U44. Since no substrate is present during the photochemistry, cross-links from this region must be the result of an alternate structure. One such structure, which often appears in computer secondary structure folds, is the

Figure 14. *Intramolecular Ribozyme Cross-link Analysis*

Reverse transcriptase analysis of intramolecular cross-link sites on the ribozyme. R and RM lanes are analysis of unreacted and monoadducted ribozyme. Lanes 1 through 8 analyze intermolecular cross-links isolated during the second round purification, see Figure 8. Lane 8r was isolated during the first round of purification, see Figure 7. Sequencing lanes are presented as their representative complementary RNA position rather than the cDNA sequence itself. Major cross-link stops are indicated by the labelling on the right side of the figure. The cross-linking pattern was identical for isolated intramolecular cross-links from both the Ribozyme alone and Ribozyme + Substrate experiments, see Figure 9.

38

Figure 15. Possible Intramolecular Cross-links in the Ribozyme

Nucleotides highlighted in bold type are psoralen reactive sites. Cross-links are represented by the ~ symbol.

formation of a hairpin within ribozyme nucleotides 37-47 (Fig. 15). This hairpin allows the possibility for several cross-links, occurring between U39, U41, U43, and U44, as shown in Figure 15. Formation of this hairpin disrupts helices 3 and 4. Another possible cross-link is between U39 and C27.

Other significant stops include U33, C27 and C6. For U33 there are a few different ways in which secondary structures can be drawn which allow a cross-link to occur between U33 and some location towards the 5' end of the molecule, especially since the 5' end of the ribozyme is adenine rich. Several structures suggested by the secondary folding

program, though there is no real evidence to confirm the existence of any of these structures.

The nature of the reverse transcriptase analysis method suggests that absolute stops occur at the 3' end of a cross-link and that the invisible end of the cross-link must be towards the 5' end of the molecule. Given this fact, there is no conventional way to explain cross-links which have their 3' ends at C27 and especially at C6, the final pyrimidine at the 5' end. All explanations would be wild guesses at best and therefore none will be given.

(f) Site-directed Mutagenesis

Site-directed mutagenesis was used to answer two key questions. Which of the conformations, determined by psoralen cross-linking, is the one responsible for cleavage activity and what is the role of each individual nucleotide within the active conformation?

The original approach to answering these questions was to use the classic mutant revertant method. For each true base pair, mutating only one nucleotide should disrupt pairing and destroy activity. The activity could then be restored by mutating the second nucleotide to return complementarity. The cross-linking data suggests several possible conformations, each of which has different base pairings. By testing different mutant revertant combinations we should be able to select out which of the base pairs are responsible for catalytic activity. While this approach works well with standard double helical regions, the added complexity of internal loop systems made it unworkable. Each of the nucleotides in the internal loop regions is capable of two different types of interactions. The first is tertiary interactions, such as those identified by the cross-linking experiments. The second is secondary interactions with nucleotide across the internal loop. Any change in the nucleotide sequence which alters the possible tertiary interactions also alters the possible secondary interactions. For example, one possible mutant revertant pair is to change ribozyme A40 to U40, and substrate U9 to A9. From the stand point of the

tertiary interaction this could work. However, there is now a run of three uridines directly across from three adenines within the ribozyme internal loop. Needless to say, such complementarity results in the collapsing of a large part of the internal loop into a standard double helix, negating any possibility for a tertiary interaction. In addition, there are likely to be special properties created by the specific sequences used in these internal loop regions. It is probably no coincidence that both internal loops contain the sequence AGAA. The fact that RNA is capable of making a variety of stable non-Watson-Crick pairings complicates matters further (Saenger, 1984). Wimberly and Tinoco (personal communication) have shown that changing the sequence of internal loop E of *Xenopus laevis* 5S RNA by one nucleotide collapses the loop. This collapsed loop contains such unconventional pairings as GA, AA, and UU.

In order to reduce the problems of interpretation that mutant revertant experiments produce a different, though more limited, mutation strategy was tried. Instead of trying to make mutant revertant pairs we approached the problem by making single mutations that, depending on which conformation is the active conformation, might have positive, neutral or negative effects on activity. Because the possibility of point mutants to create interactions within the internal loop still exists, only positive or neutral results can provide any clear evidence in favor of one of the four potential conformations suggested by cross-linking (Fig. 11, conformations A-D). Seven different mutations were made in the region of the ribozyme involved in the tertiary interactions: A40 to G, A40 to U, U41 to C, A42 to G, A42 to U, U43 to C, and U44 to C. The experimental results are summarized in table 1. The majority of these mutants showed little or no activity. The only mutant having full activity was C41. Because 6 out of 7 of these mutants gave negative results, they in themselves do not give us enough information to help distinguish between the three possible active conformations. Likewise, mutant C41 which gave a non-negative result cannot in and of itself provide enough information to allow for us to discriminate between the three conformations.

Hampel *et al.* (1990) found that of the four nucleotides that comprise the substrate side of one of the internal loops, AGUC, only one nucleotide, A7, could be varied with no effect on the ribozyme activity. As part of the attempt to produce mutant revertant experiments U9 was mutated to C and A. If a base pair with U9 is an essential part of any tertiary interaction changing it should destroy activity. The results of these two mutations, listed in table 1, show that activity is greatly diminished but not completely eliminated. This data suggests that either these mutants are capable of alternate noncanonical base pairings or that reducing the number of base pairs from 3 to 2 is still sufficient for activity.

Table 1: Mutant Results

	Mutation	Activity (percent of w.t.)
	w.t.	100
Substrate	A9	7
	C9	15
Ribozyme	G40	2
	U40	<1
	C41	100
	G42	3
	U42	3
	C43	25
	C44	3
	Del44	NC
	C10	NC
	C23	58
Helix 4	R1del (Del28)	NC
	R1ext/del	NC
	R1ext	100
	R2ext	<1

NC = No detectable cleavage. Rates for these reactions are limited to less than 0.01 percent of wild type.

Several other mutation experiments, listed in table 1, were performed to answer questions about other aspects of the molecule's structure. To test the importance of the

internal loop length a deletion mutant was made at position 44 in the ribozyme. This deletion shortens one half of the internal loop while maintaining virtually all of the sequence required for the different tertiary interactions. The result of this deletion was the complete loss of activity. Because this mutation occurs within the internal loop it suffers from the same complications as the previously described mutants. However, the fact that it completely destroyed activity, whereas all except one of the other point mutants retained some residual activity, suggests that it has altered a more fundamental aspect of the system rather than just altering the equilibrium from tertiary to secondary interaction. One possibility is that the shortened loop length critically alters the geometry of the helix created by the tertiary interactions in relation to the cleavage site.

Two mutations were directed towards addressing the role of the non-cross-linkable halves of the two internal loops. The fact that they both contain the identical sequence, AGAA, suggests that they may serve similar functions. One possibility is that the AGAA sequence produces particularly stable, purine rich, single stranded stack, helping to promote the tertiary interaction over the secondary ones. To test such a hypothesis, the two sequences were mutated in an identical manner, G was replaced by C to give ACAA. This mutation does not create any obvious base pair interactions within either internal loop and our guess would be that they would affect activity in a similar manner. The actual results, listed in table 1, turned out to be dramatically different for the two mutants. While the mutation between helices 3 and 4 maintained more than half of its wild type activity, the other mutant was completely dead. This again suggests more than just loop collapse. The AGAA sequence across from the cleavage site probably does more than just stand out of the way and may be involved in some presently undetermined tertiary interaction.

The final set of mutants, also listed in table 1, were made to address questions about the conformation of stem 4 within the cleaving structure. The secondary structure prediction programs produced two energetically similar conformations for helix 4, R1 with 5 base pairs and R2, shown in Figure 16A. In addition, the cross-linking data (Fig. 11,

conformation D) suggests that there are times when this end of the molecule has no secondary structure at all. A single mutant revertant experiment by Hampel *et al.* (1990) led them to predict that helix 4 conformation R1 was likely to be the active structure. In order to confirm their results and to address the two part question, is the bulged A essential and does the active conformation have a 5 base pair/internal bulge hairpin or just a 3 base pair hairpin, the mutants in Figure 16B were made. The two different hairpin configurations, R1 and R2, were firmly established by extending each of them by four base pairs. The bulged A/base pairing question was addressed by deleting the A from both the native and extended hairpin structures. The result was that the only active structure was R1ext. Both of the deletion mutants were completely inactive, and R2ext showed only the slightest amount of residual activity. The data demonstrates that the three base pair hairpin is formed within the active structure and that the AC to GU part of the stem is not. The real difference between the active R1 with the AC/GU unbase paired and inactive R2 is the length of one side of the internal loop. In R2 it is 5 bases long, AGAAA, and in R1 it is 7, AGAAACA, suggesting that there is a minimum length requirement for this loop.

Even though the R1ext structure and the native RNA structure have identical rates of cleavage, the overall cleavage yield is higher for R1ext, 93% versus 86% after 60 minutes of incubation. The R1ext structure reduces the population of RNA complexes that may be trapped in an inactive, alternate hairpin conformation, thus improving the overall yield.

A **Helix 4 conformations**

B **Helix 4 mutations**

Figure 16. Helix 4 Conformations and Mutagenesis

A: The active conformation of helix 4 is indicated by the *. B: Nucleotides added to extend the length of helix 4 are highlighted in bold type.

Discussion:

Without extensive X-ray crystal or NMR data no exact structure can be defined for the self-cleaving sequence we have studied. What our work has been able to do is define most of the important themes and interactions within the molecule, providing the foundation for more detailed studies. Figure 17 shows a summary of the results from our lab and others. The molecule consists of four secondary helical regions centered around a hinge, and is otherwise divided by internal loop regions. These two single stranded regions come together to form a tertiary interaction. The presence of helices 1 through 4 have been substantiated by mutagenesis. Our data suggests that Helix 4, in the active conformation, is only a 3 base pair helix and not the 5 base pair with a bulged A helix suggested by Hampel *et al.* (1990).

It is clear from the cross-linking data that there are tertiary interactions which occur between the two internal loop domains and that several different conformations can occur. These tertiary interactions are in equilibrium with each other and with secondary interactions that occur within the internal loop regions. Of the four conformations shown in Figure 17 only one is likely to be responsible for the cleavage activity.

Two of these conformations can be excluded on the basis of mutagenesis data. Conformation B contains four potential base pairs. While this conformation may provide extra stability it is not likely to be the active cleaving conformation. Mutagenesis experiments by Hampel *et al.* (1990) have shown that the adenosine on the 5' side of the cleavage site is completely variable suggesting that this base moiety is not base paired and flipped out into solution. The presence of the uridine across from the adenosine, in conformation B, promotes standard A-form base pair structure formation preventing the possibility for an unusual conformation for the base, sugar and phosphate groups of the adenosine. The existence of this A-U base is substantiated by the cross-link between U44 of the ribozyme and C6 of the substrate, which also demonstrates that the helix formed by the tertiary interaction can stack onto helix 2 allowing for very little disruption of the helix

Structure Summary

Four different tertiary conformations

Figure 17. Summary of Structural Results

backbone on the substrate side. The other three conformations do not suggest any easy interactions with this adenosine.

Conformation D is the other conformation which can be easily ruled out. Conformation D requires the complete disruption of helix 4. Extensions of helix 4 should prevent or at least reduce the frequency of disruption of this helix, and inhibit the formation of conformation D. Such a mutation should have an adverse effect on the rate of cleavage, yet no reduction in rate was seen.

Elimination of conformations B and D reduces the possibilities to two. None of the mutants presented here provide enough information to discriminate between the two conformations, A and C, and barring some great fortune, standard mutagenesis alone cannot determine the active conformation. Another approach to determining the structure of this molecule is being performed by the Bruening lab at U.C. Davis (personal communication). Their approach, which uses a unimolecular system, has been to walk through the structure making each site abasic, and determining the impact of this modification on activity. This method has the advantage of not creating new interaction possibilities upon modification. What they discovered is that only four nucleotides can, as individuals, eliminate activity, they are U43, U44, and A45 on the ribozyme, and G8 on the substrate. These four nucleotides are all in the internal loop regions and map nearly exactly to the nucleotides that comprise the tertiary interaction within conformation A. This data combined with ours indicates that this conformation is the tertiary interaction present within the active structure. From an energy standpoint conformation A is slightly more stable than conformation C because it has a G-C base pair rather than G-U. This G-C pair is absolutely essential. The other two base pairs that make up the tertiary interaction were able to be eliminated and still allow the reaction to occur, albeit at a reduced rate. If both of these base pairs were eliminated simultaneously, reducing the tertiary interaction to one base pair, the reaction would probably not occur.

The use of a unimolecular system by the Bruening lab greatly improves the ability of the molecule to undergo ligation. Changing positions 9 and 10 of the substrate to abasic sites prevents ligation. This fact suggests that although U9 and C10 may not be required for cleavage it is essential to form base pairs to these two nucleotides to promote ligation.

Between the three known sequences (Figure 3) there is some variability in the nucleotides that are involved in tertiary base pairs in conformation A. For both (-)SArMV and (-)SCYMV-S1 the sequence flanking the cleavage site changes from AGUC to AGUA. The change of C to A converts the putative C-U base pair into a bonafide Watson-Crick A-U base pair which will strengthen the interaction. The other change is in the ribozyme UAC sequence. In (-)SCYMV-S1 this sequence becomes UAU, which changes a G-C base pair to G-U. This change does not greatly reduce the overall stability of the tertiary interaction and is compensated by the new C-U to A-U base change.

When one looks at secondary representations of this molecule it can be difficult to understand how the two internal loop sequences can come together to form base pairs without introducing large steric conflicts. This is especially true for the tertiary interaction of conformation A, which has only one single stranded nucleotide bridging the junction between helix 2 and the tertiary interaction helix. A first step to understanding how easy or difficult this interaction may be to make, for any of the four conformations, is to calculate the helical rotations within helices 2 and 3. Figure 18 shows a simplified representation of helices 2 and 3. At the hinge, the two strands that are involved in the tertiary interaction are diametrically opposed to each other. As we move forward along the sequence backbones helix 2 rotates 155 degrees and helix 3 rotates 124 degrees. These rotations reduce the distance between the two helical backbones from twice the width of the helix, about 36 Å, to about 20 Å. This distance is only slightly larger than the average helical width, thus even if one wanted to base pair the first nucleotides that emerge from the two helices, and they were receptive to base pairing, there would only be a slight distortion in the sugar-phosphate backbone. As it is, A7 on the substrate is unbase paired and allows additional

flexibility for the two internal loop sequences to come together in any of the four conformations suggested by the cross-linking results.

Figure 18. *Simplified Representation of Helices 2 and 3.*

How the two internal loops are able to come together to form the tertiary interaction within conformation A is shown schematically in Figure 19. In the end on view, A7 of the substrate is shown swung out into solution, as has been suggested by models for altering the sugar-phosphate backbone into a reactive conformation. The side view shows how the two sugar-phosphate backbones flow easily from helices 2 and 3 to come together to form the tertiary interaction.

Figure 19. Simplified Representation of Helical Progression from Helices 2 and 3 to the Tertiary Helix.

A view of the overall sugar-phosphate backbone and hydrogen bonding pattern is shown in Figure 20. This picture was produced by entering the helical and backbone constraints into the program DSPACE followed by computer and byhand rendering. Again, we see that the helices bend easily around the hinge, allowing for the tertiary interaction to occur. The molecule modelled is unimolecular with the 3' end of the substrate joined to the 5' end of the ribozyme. This conformation produces an overall shape that looks very reminiscent of a twisted paperclip. This observation that the backbone followed a path similar to a paperclip was originally made by members of George Bruening's lab. At the center of the "paperclip" is the tertiary interaction and cleavage/ligation site. This interaction is flanked on the ends by the four helices and on the outside by the two unbase paired internal loop sequences.

If the molecule was comprised of only the four helical interactions and the one tertiary interaction one would expect some variability in the two single stranded sequences that are not involved in any obvious base pairs. To the contrary, we find that these internal loops sequences are highly conserved between the three known sequences (Figure 3). This lack of variability suggests that these sequences play an important role in the structure. The easiest conclusion to draw is that these sequences were chosen because they minimize any possible secondary interactions within the internal loops, promoting the formation of the tertiary interaction. Other possible roles are not so obvious. From the literature we know that base triples and other unconventional interactions are routinely found in RNA structures (Saenger, 1984). The view in Figure 20 reveals that the two sequences not involved in the base paired tertiary interaction, shown in light shading, are still in very close proximity of the other parts of the molecule, especially the tertiary region, and could very well be involved in noncanonical interactions with various phosphates, sugars and base functional groups. The mutagenesis of G10 of the ribozyme hints at such possibilities (see mutagenesis results).

Figure 20. Model of "Paperclip" RNA in Cleaving Conformation A.

Backbone and base pair trace of the "paperclip" RNA calculated by DSPACE and rendered by computer and byhand methods. The molecule is shown in unimolecular form, with the 3' end of the substrate joined to the 5' end of the ribozyme. Lightly shaded areas are conserved internal loop regions potentially involved in non-Watson-Crick tertiary interactions.

Two major insertions have been made into the ribozyme sequence which surprisingly do not destroy the cleavage activity (Haseloff and Gerlach, 1989). These sequences and how they are likely to insert in the ribozyme are shown in Figure 21. These insertions disrupt the helical regions and the possible tertiary interactions only to a minor extent. Insertion D71 extends the size of the one of the internal loop sequences. This insertion does not create any steric problems since this loop is on the outside of the structure. It does suggest that while there may be a minimum length for this internal loop sequence there is no reasonable maximum constraint. The second insertion, D86, enlarges the hairpin loop of the ribozyme. The insertion does disrupt one of the three base pairs of helix 4 which may reduce the cleavage rate but the additional sequence does not affect the catalytic structure in a similar manner as the R1ext helix 4 extension did not affect catalysis.

Figure 21. *Structure of Two Catalytically Active Insertion Mutants*

Inserted nucleotides are highlighted in bold type.

"Hammerhead" versus "Paperclip"

Upon first examination, the "hammerhead" and "paperclip" motifs appear to be as different as the common objects they are named after (Figure 22). The "hammerhead" is a three stem junction with a curious tertiary blob in the center. The "paperclip" is a hinged set of helices and internal loops which come together to give parallel helices connected at the center. This molecule is more akin to a Holiday junction than a three stem junction. However, as one looks more carefully at the basic motifs of these two different self-cleavers, more commonality can be found. In both sequences the cleavage products are the same suggesting a similar mechanism and a similar conformation for the nucleotide on the 3' side of cleavage and its 3' phosphate. In both ribozymes unbase paired nucleotide, can be varied, and is likely to be projected into solution. The choice of what this nucleotide is is probably guided by which nucleotide is least likely to form a secondary interaction. To either side of this nucleotide the substrate is bound in helices, helices 1 and 3 in the "hammerhead" and helix 2 and the tertiary helix in the "paperclip". In both cases the unbase paired, conserved sequence adjacent to the cleavage site is GAA.

Figure 22. "Hammerhead" versus "Paperclip"

In order to show these similarities the two molecules have been redrawn in a more folded form shown in Figure 23. Figures A and C show the related cores of these two molecules. The two helices surrounding the cleavage site have been drawn parallel but not stacking on each other. The models of the “paperclip” suggest that this is so, as does the fact that there are many examples of bulged nucleotide helices which stack that do not promote autolytic cleavage. Helix 2 of the “hammerhead” has been oriented parallel also and correlates to the position of helix 1 of the “paperclip”. The conserved internal loop sequence in the “hammerhead”, CUGAUGA, overlaps the center of the molecule (Figure 23B), making a variety of tertiary contacts with the GAA sequence and helical moieties possible. The AGAAACA sequence from the “hairpin” can be drawn in a similar fashion (Figure 23D) and potentially makes similar tertiary contacts. The most likely role for these nonstandard tertiary interactions is to orient the two helices bounding the cleavage site into a specific structure promoting autolysis.

Why a “paperclip” and not a just another “hammerhead”? Their catalytic cores appear to be very similar, suggesting that one could be replaced by the other. The “paperclip”, however, has some additional conserved sequence and very different helical linkages. These differences and their reasons for existing are dictated by the different roles of the positive and negative strands in the rolling circle replication mechanism. The efficiency of this mechanism is improved by having the intermediate, negative strand undergo ligation in high yield to produce the positive strand template. The positive strand does not require efficient ligation. Its mature form, which is packaged and disseminated, is linear.

It is well established that the “paperclip” containing negative strand of STobRV can efficiently perform the religation of cleaved monomers into circles, whereas the “hammerhead” containing positive strand of STobRV cannot. While the rule of microscopic reversibility says that the conformation which allows cleavage to occur must also allow ligation, it does not specify where the equilibrium between cleaved and ligated forms must lie. For the “hammerhead” structure this equilibrium lies far to the side of

Figure 23. *Common Motifs Shared by the "Hammerhead" and "Paperclip" Structures*

A & C: The helical and sequential cores shared by the "hammerhead" and "paperclip" molecules. B & D: Full representations of the two molecules. Highlighted in bold are the conserved loop sequences which are likely to be involved in tertiary interactions with the cleaving core. The cleavage site is indicated by a *. The "paperclip" sequence is shown in conformation A.

cleavage and by analogy one would expect the conformation of the "paperclip" which shares so much similarity with the "hammerhead" to also have an equilibrium far toward the side of cleavage. The "paperclip", however, does have other conformations which readily form. One conformation, conformation B, was originally dismissed because the structure forms a base pair with the A on the 3' side of the cleavage site and promotes coaxial stacking with helix 2. While this conformation is unlikely to promote a high level of cleavage it is ideal for ligation. Hampel *et al.* (1990) have shown that the nucleotide on the 3' side of the cleavage site is totally variable and does not affect cleavage activity, they did not, however, examine the affect of these mutations on ligation. Phylogenetic data from the three known structures argues that this nucleotide and the sequence which binds to it in conformation B cannot be varied, suggesting that this conformation with it's specific base pairing is biologically essential. Thus, it appears that the "paperclip" utilizes at least two of it's conformations, one promoting cleavage and the other ligation. The "hammerhead" cannot do this and therefore is replaced by a more flexible sequence.

References:

- Arnott, S., Hukins, D. W. L., Dover, S. D., Fuller, W., and Hudgson, A. R. (1973). *J. Mol. Biol.* **81**, 107-122.
- Bruening, G. (1989). "Compilation of Self-cleaving Sequences from Plant Virus Satellite RNAs and Other Sources." *Methods Enzymol.* **180**(RNA Processing, Pt. A), 546-558.
- Bruening, G., Gould, A. R., Murphy, P. J., and Symons, R. (1982). "Oligomers of Avocado Sunblotch Viroid are found in Infected Avocado Leaves." *FEBS Lett.* **148**(1), 71-78.
- Buzayan, J. M., Gerlach, W. L., and Bruening, G. (1986). "Non-enzymatic cleavage and Ligation of RNAs Complementary to a Plant Virus Satellite RNA." *Nature* **323**, 349-353.
- Caviani Pease, A. and Wemmer, D. E. (1990). "Characterization of the Secondary Structure and Melting of a Self-Cleaving RNA Hammerhead Domain by ^1H NMR Spectroscopy." *Biochemistry* **29**, 9039-9046.
- Cimino, G. D., Gamper, H. B., Isaacs, S. T., and Hearst, J. E. (1985). "Psoralens as Photoactive Probes of Nucleic Acid Structure and Function: Organic chemistry, photochemistry, and biochemistry." *Annu. Rev. Biochem.* **54**, 1151-1193.
- Epstein, L. M. and Gall, J. G. (1987). "Self-Cleaving Transcripts of Satellite DNA from the Newt." *Cell* **48**, 535-543.
- Ericson, G. and Wollenzien, P. (1988). "Use of Reverse Transcription to Determine the Exact Locations of Psoralen Photochemical Crosslinks in RNA." *Anal. Biochem.* **174**(1), 215-223.
- Ericson, G. and Wollenzien, P. (1989). "An RNA Secondary Structure Switch between the Active and Inactive Conformations of the *Escherichia coli* 30S Ribosomal Subunit." *J. Biol. Chem.* **264**(1), 540-545.
- Fedor, M. J. and Uhlenbeck, O. C. (1990). "Substrate Sequence Effects on "Hammerhead" RNA Catalytic Efficiency." *Proc. Natl. Acad. Sci. USA* **87**, 1668-1672.
- Feldstein, P. A., Buzayan, J. M., and Bruening, G. (1989). "Two Sequences Participating in the Autolytic Processing of Satellite Tobacco Ringspot Virus complementary RNA." *Gene* **82**(1), 53-61.
- Forster, A. C., Davies, C., Hutchins, C. J., and Symons, R. H. (1990). "Characterization of Self-Cleavage of Viroid and Virusoid RNAs." *Methods Enzymol.* **181**(RNA Processing, Pt. B), 583-607.

- Hampel, A. and Tritz, R. (1989). "RNA Catalytic Properties of the Minimum (-)sTRSV Sequence." *Biochemistry* **28**, 4929-4933.
- Hampel, A., Tritz, R., Hicks, M., and Cruz, P. (1990). "'Hairpin' Catalytic RNA Model: Evidence for Helices and Sequence Requirement for Substrate RNA." *N.A.R.* **18**(2), 299-304.
- Haseloff, J. and Gerlach, W. L. (1989). "Sequences Required for Self-Catalysed Cleavage of the Satellite RNA of Tobacco Ringspot Virus." *Gene* **82**, 43-52.
- Heus, H. A. and Pardi, A. (1991). "Nuclear Magnetic Resonance Studies of the Hammerhead Ribozyme Domain." *J. Mol. Biol.* **217**, 113-124.
- Hubbard, J. and Hearst, H. (1991). *J. Mol. Biol.* in press.
- Jaeger, J. A., Turner, D. H., and Zuker, M. (1989a). "Improved Predictions of Secondary Structures for RNA." *Proc. Natl. Acad. Sci. USA* **86**, 7706-7710.
- Jaeger, J. A., Turner, D. H., and Zuker, M. (1989b). "Predicting Optimal and Suboptimal Secondary Structure for RNA in "Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences"." *Meth. Enzymol.* **183**, 281-306.
- Kaper, J. M., Tousignant, M. E., and Steger, G. (1988). "Nucleotide Sequence Predicts Circularity and Self-Cleavage of 300-Ribonucleotide Satellite of Arabis Mosaic Virus." *Biochem. Biophys. Res. Com.* **154**(1), 318-325.
- Kiefer, M. C., Daubert, S. C., Schneider, I. R., and Bruening, G. (1982). "Multimeric Forms fo Satellite of Tobacco Ringspot Virus RNA." *Virology* **121**(2), 262-273.
- Lipson, S. E., Cimino, G. D., and Hearst, J. E. (1988). "Structure of M1 RNA as Determined by Psoralen Cross-Linking." *Biochemistry* **27**, 570-575.
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Maser, R. L. and Calvet, J. P. (1989). "U3 Small Nuclear RNA can be Psoralen-Cross-Linked *in vivo* to the 5' External Transcribed Spacer of Pre-Ribosomal-RNA." *Proc. Natl. Acad. Sci. USA* **86**(17), 6523-6527.
- Mei, H.-Y., Kaaret, T. W., and Bruice, T. C. (1989). "A Computational Approach to the Mechanism of Self-Cleavage of Hammerhead RNA." *Proc. Natl. Acad. Sci. USA* **86**, 9727-9731.
- Milligan, J. F., Groebe, D. R., Witherhall, G. W., and Uhlenbeck, O. C. (1987). "Oligoribonucleotide Synthesis Using T7 RNA Polymerase and Synthetic DNA Templates." *NAR* **15**, 8783-8798.
- Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R., and Bruening, G. (1986). "Autolytic Processing of Dimeric Plant Virus Satellite RNA." *Science* **231**, 1577-1580.

- Rubino, L., Tousignant, M. E., Steger, G., and Kaper, J. M. (1990). "Nucleotide Sequence and Structural Analysis of 2 Satellite RNA's Associated with Chicory Yellow Mottle Virus." *J. Gen. Virology* **79**, 1897-1903.
- Ruffner, D. E., Stormo, G. D., and Uhlenbeck, O. C. (1990). "Sequence Requirements of the Hammerhead RNA Self-Cleavage Reaction." *Biochemistry* **29**, 10695-10702.
- Saenger, W. (1984). Principles of Nucleic Acid Structure. New York: Springer-Verlag. 556 pp.
- Uhlenbeck, O. C. (1987). "A Small Catalytic Oligoribonucleotide." *Nature* **328**, 596-600.
- Youvan, D. C. and Hearst, J. E. (1982). "Sequencing Psoralen Photochemically Reactive Sites in *Escherichia coli* 16S rRNA." *Anal. biochem.* **119**, 86-89.
- Zuker, M. (1989). "On Finding All Suboptimal Foldings of an RNA Molecule." *Science* **244**, 48-52.

Chapter 2

Part One: Development of the Transcript Folding Assay and Application to *E. Coli* RNA Polymerase

Introduction:

Transcription is the most important level of control of gene expression. The process of transcription can be divided into four stages (Chamberlin, 1974): promoter binding, initiation of transcription, elongation of the transcript, and termination. Mechanistic details of all of these processes are becoming available, and it is clear that cellular regulatory mechanisms act at all four stages. Our work concerns the structure and dynamics of the ternary polymerase•DNA•RNA complex during elongation.

The formation of structure in the nascent RNA chain is an essential step in pausing, antitermination, and termination of transcription by *E. coli* RNA polymerase. These processes are believed to require the formation of an RNA hairpin near the 3' end of the transcript. The prevailing model for "simple", rho factor independent transcription termination and for pausing requires that an RNA hairpin disrupt the RNA–DNA hybrid helix within the transcription bubble (Farnham and Platt, 1980; von Hippel *et al.*, 1984; Yager and von Hippel, 1987; Yager and von Hippel, 1989). RNA structure formation during transcription is required for attenuation of transcription in the *E. coli trp* operon (Landick and Yanofsky, 1987). The action of several transcription regulation factors such as NusA, proteins N and Q, and rho are affected by RNA structure (Brennan *et al.*, 1987;

Faus *et al.*, 1988; Landick and Yanofsky, 1987; Yang and Roberts, 1989). These proteins have been shown to affect pausing and termination of transcription, but in some cases it is not clear whether they interact with the nascent transcript, the polymerase itself, or both. Specific folding patterns in the nascent RNA have been proposed to direct splice site selection (Eperon *et al.*, 1988) and to mediate Tat protein activation of HIV-1 gene expression (Berkhout *et al.*, 1989), and to affect RNA-RNA interactions responsible for the control of plasmid ColE1 replication. Transient RNA structure could also affect RNA transport, assembly of protein-RNA complexes, or other aspects of RNA metabolism. Similar considerations may be important in protein folding during translation or membrane translocation (Rothman, 1989).

The order in which subsets of RNA sequence become available for folding may kinetically determine the final transcript structure, especially for large RNA molecules in which many conformations may be thermodynamically equivalent but once formed are too stable to interconvert. It has been shown experimentally that some RNA structures are present only transiently during RNA synthesis by Q β replicase (Kramer and Mills, 1981). Possible examples of kinetically dependent structures important *in vivo* are long multiple intron mRNA's and the *E. coli* 16S ribosomal RNA. Ericson and Wollenzein (1989) have identified conformational differences between active and inactive forms of 16S RNA. The active conformation appears to be favored by kinetically driven structure formation, meaning that one of a large population of energetically similar conformations is formed preferentially during transcription. The active and inactive conformations are otherwise metastable and require heating to interconvert.

Knowledge of the detailed structure of newly formed RNA and of when it becomes available for intramolecular interactions could provide insight into the molecular mechanisms of processes such as termination and pausing, which require hairpin formation to occur within a few nucleotides of the site of polymerization. Studying the effects of

antitermination factors on transient structure could help determine whether their effects are mediated through the transcript or through protein-protein interactions.

RNA structure formation during transcription antitermination has been studied using *E. coli trp* operon mutations and RNase T₁ probing of paused transcription complexes. The transcript forms structure upstream from an RNA polymerase paused at the base of the 1:2 hairpin. The exact structure of the presumed RNA hairpin, how much of the hairpin forms, and whether it interacts with or disrupts the RNA-DNA hybrid is not known (Landick and Yanofsky, 1987).

Several groups have examined the contacts and accessibility of the nascent RNA chain. Hanna and Meares (1983) have used cleavable photoaffinity labeling techniques to show that the 5' end of a growing RNA chain labels the DNA coding strand for 12 to 14 base pairs upstream from the point of synthesis on the T7 A1 promoter. The leading end of the transcript becomes accessible to DTT as transcript length increases from 12 to 14 nt (Bernhard and Meares, 1986). RNA chains up to 94 nucleotides in length can contact the β and β' subunits of polymerase. The existence of these RNA-DNA and RNA-protein interactions does not indicate that the RNA does not interact with itself, especially as photoaffinity labeling yields are typically low (Stackhouse and Meares, 1988). Additionally, the RNA-DNA contacts identified are those present in the initiation complex. The stability of these complexes is limited because of abortive initiation, and the σ subunit is still often bound. The RNA-DNA contacts may be significantly different in the elongation complex; photoaffinity labels which can be incorporated into the body of an RNA chain are now available to probe elongation complex interactions (Hanna *et al.*, 1989).

Kumar and Krakow (1975) performed RNase digest experiments on *Azotobacter vinelandii* RNA polymerase transcription complexes. The amount of RNA protected from RNase T₁ digestion was found to be between 8 and 14 nucleotides. The photoaffinity and digestion data are consistent with the existence of an RNA-DNA hybrid of length 12 ± 2

base pairs exists in the transcription bubble, though direct evidence for RNA–DNA base pairing was not obtained.

Our experiments address the exact point at which an RNA transcript becomes free to form structure with itself, and whether this structure is capable of forming immediately adjacent to the RNA–DNA hybrid or interacting with the hybrid. To this end we have developed a simple, nonintrusive assay for structure formation, using the RNA transcript itself as a probe. By using an RNA molecule which autolytically cleaves when it forms structure we can identify exactly how far transcription must proceed before the RNA transcript is free enough to form the self–cleaving structure. This assay introduces no artifacts due to the presence of exogenous probes and accesses the elongation complex rather than an initiation complex. It also provides a simple method of assessing the effects of accessory proteins such as NusA on the transcript structure.

RNA capable of autocatalytic self cleavage is common among certain viroids and virusoids and other RNA molecules (Forster and Symons, 1987; Prody *et al.*, 1986; Uhlenbeck, 1987), reviewed in Keese and Symons (1987). This “hammerhead” structure requires the presence of 13 consensus sequence nucleotides, interacting to form an unknown tertiary structure linked by three helical stems. The cleavage reaction is site–specific and yields 2′,3′-cyclic phosphate and 5′-OH termini. Uhlenbeck (1987) has measured the rate of cleavage for “hammerheads” constructed from two separate molecules. The substrate RNA’s have half lives as short as 2 minutes, and unimolecular cleavage rates are likely to be even faster. The cleavage reaction provides an internal assay for RNA secondary and tertiary structure formation since only a fully formed “hammerhead” is capable of cleavage. At the salt concentrations used in transcription, the cleavage reaction is absolutely dependent on Mg^{2+} or other divalent cations, so cleavage can be rapidly and conveniently prevented by the addition of EDTA.

We find that 9–13 nt of RNA beyond the final stem of the “hammerhead” structure are required for cleavage, indicating that this length of RNA is restrained by interactions in the

ternary complex. These results are consistent with the proposed length of the RNA–DNA hybrid, and they show that once the transcript is free of the hybrid its folding is not constrained by other interactions within the ternary complex. The self–cleaving transcript experiment is very simple and straightforward, and it is readily applicable to other polymerase systems including complexes modified by accessory transcription factors.

Materials and Methods:

(a) Materials

Purified *E. coli* RNA polymerase holoenzyme and NusA protein were generously provided by Prof. Michael Chamberlin (University of California, Berkeley). Purified T7 RNA polymerase was purchased from Pharmacia. Ribonucleoside triphosphates, 3'-deoxyribonucleoside triphosphates (3'-dNTP's), 3'-O-methylnucleoside triphosphates (3'-O-MeNTP's), and ribonucleases were purchased from Pharmacia. Ribonuclease inhibitor (RNasin, human placenta) was from Promega Biotec or Pharmacia. Adenosine 5'-(1-*O*-thio)triphosphate ([α -thio]ATP) was obtained from New England Nuclear. [γ -³²P]Adenosine 5'-triphosphate and T4 DNA ligase were from Amersham. Cloned T4 polynucleotide kinase was supplied by New England Biolabs. Guanylyl(3',5')guanosine (GpG) and heparin were purchased from Sigma. Electrophoresis grade acrylamide, bis(acrylamide), ammonium persulfate, and TEMED were obtained from Bio-Rad. Ultra pure urea was from Schwarz/Mann Biotech. Diethylpyrocarbonate-treated water, silanized Eppendorf tubes (5% dichlorodimethylsilane/chloroform), and deionized formamide were used throughout.

(b) DNA Templates

Deoxyoligonucleotides (45–55 nt) were synthesized on an Applied Biosystems 381A automated synthesizer, purified by gel electrophoresis on 12% polyacrylamide gels, and ligated by standard methods (Maniatis *et al.*, 1982) to give the transcription templates. The oligonucleotides were lightly labelled with ³²P for ease of purification. The ligated 145 nt products were isolated on 8% polyacrylamide denaturing gels and then electrophoresed on 8% nondenaturing gels to remove any excess single strand. The identities and double-stranded character of the templates were confirmed by restriction analysis. The templates have modified TAC18 promoters (Mulligan *et al.*, 1985) immediately followed by the self-cleaving transcript sequence, as shown in Figure.1.

Figure 1. Template and Transcript Sequences

(A) Sequence of the synthetic double stranded DNA template, containing a modified TAC 18 promoter followed by a sequence coding for the “hammerhead” structure. The “hammerhead” region is followed by a sequence of minimal complementarity to the “hammerhead” sequence. The control template contains a base pair change at +9 coding for a U instead of a G in the RNA product. This mutation destroys autolytic activity in the RNA.

(B) Sequence of the synthetic double stranded template, used in part two, where the TAC 18 promoter has been replaced by a modified T7 Class III promoter. The rest of the sequence is identical to that in (A). For the T7 studies in part two this sequence was incorporated into the polylinker region of pUC19.

(C) The full length RNA product. The “hammerhead”, shown in its proposed secondary structure, cleaves autolytically between nucleotides C₄₃ and A₄₄ when the “hammerhead” structure is fully formed. The conserved bases required for activity, as determined phylogenetically, are enclosed in boxes. Any deviation from this sequence results in the loss of activity, as is the case for the control transcript. In addition to the conserved nucleotides the three helical stems shown must form to promote self-cleavage. Constraints on the numbers and types of base pairs in these three stems are minimal though different sequences do exhibit a wide range of cleavage rates (Fedor and Uhlenbeck, 1990).

A: *E. Coli* RNA Polymerase template

B: T7 RNA Polymerase template

C: Folded RNA Transcript

Figure 1. Template and Transcript Sequences

(c) *In Vitro Transcription Reactions*

All transcripts were initiated with [α - 32 P]pGpG so that the 5' end of the RNA was labeled. The dinucleotide GpG was kinased in the standard buffer (Maniatis *et al.*, 1982) overnight at 4 °C using [γ - 32 P]ATP. Typical reaction mixtures included 100 μ Ci label (5000 Ci/mmol), 25 pmol GpG, and 14 U kinase in 15 μ l. The kinase reaction was heat-killed for 3 min at 65 °C and the buffer conditions were adjusted to those appropriate for RNA polymerase by adding 0.8 volumes of water (or other reaction components) and 0.2 volumes of a solution containing 1.2 M KCl, 10 mM MgCl₂, and 92 mM Tris base as described (Kahn and Hearst, 1989). Typically this mixture was then diluted about 1.5-fold by the addition of RNA polymerase and DNA as below, and further diluted 2.5-fold into the final reaction mixture.

All transcription reactions were performed at 22 °C (room temperature), except as noted, in a buffer containing 40 mM TrisOAc, pH 8.0, 6 mM Mg(OAc)₂, 20 mM KOAc, 6 mM β -ME, 0.05 mM EDTA, and 40 μ g/ml acetylated BSA. *E. coli* RNA polymerase (typically 40 nM final concentration) and a 50% molar excess of DNA template were added to the [α - 32 P]pGpG solution and the binary complex was allowed to form for 8 minutes at 22 °C, at which time the polymerase-DNA complexes were aliquoted into Eppendorf tubes containing ribonuclease inhibitor (final 2 U/ μ l), heparin (final 25 μ g/ml), and various nucleoside triphosphate and chain terminator mixtures. Final NTP and 3'-dNTP concentrations are given in the Figure legends. We found that the 3'-dNTP terminators gave more reproducible sequencing patterns and less "crosstalk" between terminators than the 3'-O-methylNTP's; 3'-O-methylNTP's are also difficult to obtain commercially. Transcripts terminated with a 3'-dNMP migrate slightly faster than the corresponding transcript with a 3' hydroxyl (see Fig. 3), which is an aid to sequencing because it allowed us to distinguish paused transcripts and specifically terminated

transcripts. The final reaction volume was 5 μ l (a minimal volume was necessary for good resolution of RNA sequencing ladders). Reactions were stopped after 30–60 min (as indicated in figure legends) by the addition of an equal volume of 100% formamide containing 20 mM EDTA and dyes. Analysis was as described below.

In experiments in which ternary complexes were disrupted and the released transcripts were allowed to self-cleave, reaction mixtures were boiled for 1.5 min, rapidly cooled to room temperature, incubated for the indicated length of time, quenched with formamide/EDTA, and analyzed as below.

(d) Transcript Analysis

Samples were concentrated in a Speed–Vac (Savant) to 5 μ l, boiled for 1.5 min, cooled on ice, and loaded directly onto 0.5 mm thick, 33 $\times$ 43 cm, 12% polyacrylamide (1:19 bis:acrylamide) gels containing 8 M urea and 0.6 $\times$ TBE (50 mM Tris, 50 mM borate, 1 mM EDTA, pH 8.3). The gels were run at 55 W (~2300 V, about 50 $^{\circ}$ C) until xylene cyanol was about 5 cm from the bottom of the gel. Minimum loading volume and high electrophoresis temperature were required for optimum resolution and complete denaturation of the transcripts. The gels were dried carefully, and end-labeled RNA was detected by autoradiography at -70 $^{\circ}$ C using Kodak XAR-5 X-ray film and Cronex Lightning Plus enhancing screens (DuPont); typical exposure times were 3–4 days.

To verify the identity of the cleavage product, enzymatic RNA sequencing was performed using standard methods as described (Kahn and Hearst, 1989; data not shown)

Results:

The principal objective of this work was obtaining information on structure formation in nascent RNA during transcription, and the interactions of the nascent RNA with the rest of the polymerase•DNA•RNA ternary elongation complex. Structure formation of the nascent RNA is affected by steric interactions within the ternary complex and by the length of the RNA–DNA hybrid, and therefore transcript structure can report on the static and dynamic structure of the ternary complex. Our approach can be applied to examining molecular processes believed to require the formation of RNA structure during elongation, such as termination and antitermination.

The autolytic cleavage reaction of the “hammerhead” RNA requires a specific secondary and tertiary structure, and therefore the self–cleavage event provides a powerful assay for structure formation. In order for cleavage to occur in a nascent transcript, folding of the “hammerhead” sequence must be free of constraints imposed by the ternary complex. By determining the transcript length necessary for cleavage, we have measured the length of RNA between the site of polymerization and the region of unhindered structure formation. This internal RNA structure assay (illustrated in Figure 2) provides a simple way to probe the ternary complex, and use of the transcript itself as an RNA structure reporter avoids artifacts potentially introduced by exogenous chemical or enzymatic probes. The “hammerhead” structure is especially useful because of its small size and well defined secondary structure.

Figure 2. *Transcription of "Hammerhead" RNA*

Schematic of transcript structure at different stages of transcription. (A) Early stage of transcription, little structure can form. (B) "Hammerhead" is fully synthesized but the final stem of the structure is still sequestered within the polymerase as part of the RNA-DNA hybrid. (C) Sequences comprising the final stem of the "hammerhead" fully emerge from the ternary complex and the "hammerhead" is able to form and self-cleave. (D) Synthesis continues after cleavage. The new 5' end is unlabeled.

These experiments require that the time scales for transcription elongation, RNA structure formation, and RNA self-cleavage be much shorter than the time required for the experiment, and disassembly of the ternary complex must be slower than all of these processes. Elongation by one base takes 20-30 ms (von Hippel *et al.*, 1984). The rate of structure formation is not well known: the relaxation time for equilibration of an RNA pseudoknot and a hairpin has been measured to be 0.1-1 sec for a process with an activation enthalpy of 42 kcal/mole (Puglisi, 1989), but the rate of processes with a lower activation energy would be expected to be much more rapid, perhaps in the millisecond range. The self-cleavage reaction rate establishes a lower limit for the rate of RNA structure formation, and the half-life for the isolated "hammerhead" used here is less than 30 sec (see below). The ternary complex is stable for hours (Levin *et al.*, 1987), and evidence presented below confirms that the ternary complex is stable for the time of the experiment. Therefore, a 30-60 min experiment provides a clear picture of the extent of the cleavage reaction at each position and the influence of the ternary complex on cleavage, and it allows us to observe cleavage events occurring with a wide range of rate constants. Varying NTP concentrations (giving slower or faster elongation) and incubation times (more or less time for cleavage or disassembly) did not substantially affect the observed set of transcripts capable of cleavage. The elongation step time may be faster than some RNA structure formation processes, but once transcription is arrested by the incorporation of a chain terminator the RNA structure has ample time to equilibrate during the experiment; when transcription is not paused or stalled there may be situations in which the relative rates of elongation and folding may affect the kinetically accessible RNA structures.

(a) Transcript Self-Cleavage in Stalled Elongation Complexes

Transcription reactions were performed in the presence of chain terminating nucleoside triphosphate analogs, usually 3'-deoxynucleoside triphosphates. A sequencing ladder of stalled complexes was created by using the four nucleotide analogs in separate reactions. A

fifth reaction using only normal triphosphates was performed to give a ladder of natural pauses and to provide a control for sequencing. Transcripts initiated at a single site from a single round of transcription were generated by using a template with the strong TAC 18 promoter and by including heparin with the triphosphates to prevent reinitiation. Only RNA molecules which had been initiated with [α - 32 P]pGpG were detected; because the labelling is at the 5' end of the initial transcript only the 5' portion of the cleaved RNA molecules are visible in these experiments. Transcription reactions were quenched with a mixture of formamide and EDTA in order to simultaneously disrupt the ternary complex and prevent further "hammerhead" cleavage by denaturing the RNA and by sequestering all divalent cations.

Two different templates were used, one coding for the "hammerhead" sequence and one control containing a single base change within the "hammerhead" consensus sequence (Fig. 1). The control template produces a transcript which does not cleave, giving a standard sequencing pattern of stalled complexes for the full length of the DNA template (Fig. 3, lanes 1-5). The transcript sequence was designed to minimize any alternate conformations which might compete with the "hammerhead" structure and inhibit cleavage. The sequence was derived from known "hammerhead" sequences and suggestions from Olke Uhlenbeck. The entire transcript (99 nt) was analyzed using the RNASS RNA folding program (Williams and Tinoco, 1986) in order to ensure that sequences in the long 3' end of the transcript would not interfere with formation of the "hammerhead" structure. An undisturbed "hammerhead" was predicted for the minimum free energy structure and for all other structures within 3.5 kcal/mole of the minimum energy (-40.5 kcal/mole at 37 °C). Visual examination of the other structures suggested that alternate conformations would be unlikely to form during transcription. The "final stem" of the "hammerhead", referred to below, is nucleotides 1-6 : 44-49; it is the last part of the "hammerhead" structure to form during transcription.

Figure 3 shows a typical autoradiogram of sequencing reactions performed on the cleaving and non-cleaving templates, using either ATP or [α -thio]ATP in the reactions on the cleaving template. The pattern of stalled complex transcripts from the "hammerhead" template is identical to the pattern from the control template for the first 42 nucleotides, after which the pattern changes as a result of transcript self-cleavage. A strong band corresponding to the cleavage product appears at position C₄₃ (the identity of the cleavage product was confirmed by enzymatic sequencing; data not shown). The patterns are again similar between A₄₄ and G₅₇, indicating that transcripts from the "hammerhead" template do not cleave in this length range. (G₅₇ indicates the transcript of length 57, whose last base is G.) Beginning at position C₅₈, there is a diminution of band intensity in the "hammerhead" lanes compared to the control template, and beyond position G₆₁ the band intensity is dramatically reduced, indicating essentially 100% cleavage of transcripts in these stalled complexes. Thus, we find that only the transcripts which extend 9 nucleotides beyond the end of the final "hammerhead" stem show any indication of cleavage, and nearly complete cleavage occurs when 13 or more nucleotides past the final "hammerhead" stem have been synthesized. This suggests that a maximum of 13 nucleotides of the transcript are unavailable for RNA-RNA base pair formation, beyond which the transcript is free to form structure.

The position of the "cutoff" for cleavage is highly reproducible. Experiments in which triphosphate or terminator concentrations were varied over a wide range (NTP's 20-400 μ M; 3'-O-MeNTP's 50-1000 μ M) gave the same cutoff position even though the overall distribution of transcript lengths was dramatically changed (data not shown). The invariance of the cutoff position indicates that differences in the fraction of polymerase that reaches a given template position are not responsible for the observed banding patterns. Note that some runoff product is visible in the "hammerhead" template lanes, and experiments described below show this is due to the presence of transcript trapped in a noncleaving conformation.

Figure 3. *Self-Cleavage of the Nascent Transcript*

Comparison of RNA sequencing ladders from transcription of the control and “hammerhead” templates shows self-cleavage of the transcript. Lanes 1-5 show transcription on the control template in the presence of the four different chain terminators and a low rNTP concentration control. Lanes 6-10 are equivalent to 1-5 except that the “hammerhead” template coding for the self-cleaving transcript is used; the cleaved product is the strong band at C₄₃. Lanes 11-15 also use the “hammerhead” template but ATP is replaced by [α -thio]ATP in the rNTP mix. The presence of a phosphorothioate at the cleavage site significantly reduces the rate of cleavage, giving a mixture of cleaved and uncleaved products. DNA templates were lightly labelled during their preparation. The G₉ $\rightarrow$ U base change is not seen in sequencing because it is off the bottom of this gel. The reaction was performed at 22 °C for 60 minutes. The triphosphate concentrations used in each lane are as follows:

Control and “hammerhead” experiments (lanes 1–10). A lanes: 3'-dATP 100 μ M; ATP 80 μ M; GTP, UTP, and CTP 450 μ M. G lanes: 3'-dGTP 100 μ M; GTP 50 μ M; ATP, UTP, and CTP 450 μ M. U lanes: 3'-dUTP 100 μ M; UTP 30 μ M; ATP, GTP, and CTP 450 μ M. C lanes: 3'-dCTP 100 μ M; CTP 50 μ M; ATP, GTP, and UTP 450 μ M. Control lanes: ATP, GTP, UTP, and CTP 50 μ M.

[α -Thio]ATP experiment (lanes 11–15). A lane: 3'-dATP 100 μ M; [α -thio]ATP 250 μ M; GTP, UTP, and CTP 450 μ M. G lane: 3'-dGTP 100 μ M; GTP 50 μ M; [α -thio]ATP, UTP, and CTP 450 μ M. U lane: 3'-dUTP 100 μ M; UTP 50 μ M; [α -thio]ATP, GTP, and CTP 450 μ M. C lane: 3'-dCTP 100 μ M; CTP 80 μ M; [α -thio]ATP, GTP, and UTP 450 μ M. Control lane: [α -thio]ATP 250 μ M, GTP 50 μ M, UTP 50 μ M, and CTP 80 μ M.

[illegible]

Figure 3

The use of [α -thio]ATP instead of ATP in the transcription reaction results in the incorporation of a phosphorothioate at the cleavage site (and other A sites), which reduces the rate of cleavage but does not substantially perturb the "hammerhead" structure (Buzayan *et al.*, 1988). Because of the reduced cleavage rate, relative to the experiment time, the phosphorothioate-containing transcripts exhibit normal elongation above G₅₇ along with cleavage product. These experiments show that the disappearance of transcripts longer than G₅₇ in the cleaving template transcriptions is due to a cleavage event rather than to enhanced pausing or other structural effects on the ternary complex. The single base change in the control transcript might potentially prevent "hammerhead" formation, allowing for normal elongation. But the fact that the phosphorothioate-containing "hammerhead" structure does not affect elongation shows that the lack of higher molecular weight transcripts is not an artifact caused simply by the formation of the self-cleaving structure.

The results above demonstrate that formation of the "hammerhead" structure and self-cleavage of the transcript are inhibited for RNA molecules within the ternary complex up to 61 nucleotides in length (corresponding to molecules including up to 12 nucleotides past the end of the final "hammerhead" stem). The control experiment of Figure 4 shows that the integrity of the ternary complex is essential for the prevention of cleavage. When the complexes are disrupted by heating, the RNA transcript is released, and the figure shows that molecules long enough to form three or more of the base pairs in the final "hammerhead" stem self-cleave, as indicated by the disappearance of all but two bands above C₄₃. The significance of the minimum three base pair requirement in the final "hammerhead" stem for self-cleavage is discussed below. Separate experiments show that cleavage is complete within 30 seconds after release (data not shown). These results also indicate that the ternary complex is stable over the course of the experiment, since released RNA would rapidly self-cleave.

Figure 4. Cleavage of Released Transcripts

Comparison of transcript cleavage before and after disruption of the ternary complex. Transcription was performed using the "hammerhead" template at 22 °C for 30 minutes, at which time the reaction volumes were split. In lanes 1–5 the reactions were immediately quenched with EDTA/formamide as usual. In lanes 6–10 the reactions were heated to 90 °C for 2 minutes to disrupt the complex and release the transcripts, quick cooled to 22 °C and incubated for 30 minutes, and then quenched. This procedure allows free transcripts of sufficient length to self-cleave. Only transcripts with two or fewer nucleotides past the cleavage site are unable to self-cleave after release, as indicated by the persistence of transcripts ending at A44 and G45. Except as noted herein the reaction conditions are the same as those for lanes 6–10 in Figure 3.

Self-cleavage of Bound and Free RNA

Figure 4

(b) Temperature Dependence

Figure 5 shows the temperature dependence of the self-cleavage reaction. As the temperature is increased from 4 °C to 50 °C the abundance of transcripts of lengths from 58 to 61 decreases until virtually all have cleaved at 50 °C. This decrease could be due to enhanced displacement of the upstream end of the RNA–DNA hybrid by the final RNA–RNA stem of the “hammerhead”, giving a larger population of molecules with fully-formed 6 base pair final stems. However, for reasons given in the Discussion, we believe instead that the enhanced cleavage is due to an increased rate of cleavage for molecules with abbreviated 3 base pair final stems. (The experiment of Fig. 4 shows that only these 3 base pairs are required for cleavage.) If this is the case, it again suggests that a minimum of 12 nucleotides, the distance from the third base pair of the stem to C₅₈, are unavailable for base pair formation. The stability of G₅₇ at 50 °C shows that this distance does vary with temperature.

Figure 5 also shows that the 99 nt runoff product and bands at several pause sites between 61 and 99 (which are visible in the control template reactions in Figure 3 as well) disappear as the reaction temperature is increased, demonstrating that these bands are probably due to a noncleaving alternate conformation. At elevated temperature, interconversion between cleaving and non–cleaving conformations is more rapid than at lower temperatures, giving efficient cleavage. Non–cleaving alternate conformations have been proposed as an explanation for the failure of some “hammerhead” sequences to cleave (Uhlenbeck, 1987). At 4 °C several intermediates of lengths between 62 and 99 are observable; at this temperature the cleavage rate even within the active conformation may be slow.

Figure 5. Temperature Dependence of Self-Cleavage Reaction

Temperature dependence of the self-cleavage reaction. The reactions were performed for 60 minutes at the following temperatures: lanes 1-5 at 4 °C, lanes 6-10 at 20 °C, lanes 11-15 at 37 °C, and lanes 16-20 at 50 °C. Except for the differences in temperature the reaction conditions are the same as those for lanes 6-10 in Figure 2.

Temperature Dependence of Self-Cleavage

Figure 5

The data above show that self-cleavage of the transcript occurs only after 12 to 13 nucleotides have been synthesized beyond the "hammerhead". When the transcript is elongated to C₅₈ the "hammerhead" sequence becomes capable of forming the minimal 3 base pair final stem, allowing cleavage to occur. Partial cleavage at 22 °C and complete cleavage at higher temperatures is observed. The results indicate that the separation between the site of polymerization and the 3 base pair final stem is at least 12 nucleotides. As the transcript is elongated to C₆₂ the "hammerhead" sequence can form the complete 6 base pair final stem, giving complete cleavage at 22 °C and higher temperatures. These data determine a maximum separation of 13 nucleotides between the site of polymerization and the end of a complete 6 base pair final stem whose structure formation appears to be unhindered.

Landick and Yanofsky (1987) and Faus *et al.* (1988) have suggested that the transcription factor NusA affects transcription by interacting with the transcript as well as the polymerase. We have performed transcription reactions similar to those above with the "hammerhead" template in the presence and absence of NusA (data not shown). The sequencing patterns were identical under the two sets of conditions, but the lack of an observable effect does not preclude the existence of NusA/RNA interactions. NusA may not interact with this particular transcript, the interaction may not disrupt the self-cleaving structure, or the time scale of the experiment may be too slow to see any effect if NusA binding is in a relatively fast equilibrium.

Discussion:

Our results clearly show that in the RNA•DNA•polymerase ternary complex the formation of intramolecular structure in the nascent RNA transcript is inhibited for the first 12 to 13 nucleotides upstream of the site of polymerization, whereas beyond 13 nucleotides RNA structure can form freely. The assay for RNA structure formation is the self-cleavage of the “hammerhead” transcript sequence as its final stem forms upon emerging from the ternary complex. This approach has advantages over intrusive structural probes such as RNases in that no artifacts are introduced due to steric interactions between the probe and the polymerase, and the functional assay for structure gives more global information than chemical protection experiments.

The self-cleavage experiments we have described are straightforward to perform, though they require practice and optimization of conditions. The experiment offers a rapid and precise measurement of the extent of RNA unavailable for folding in the transcription bubble, using readily available reagents and standard techniques. Several potential applications to other systems can be proposed. For example, polymerase accessory factors have been proposed to interact with the RNA transcript, and if any of these molecules change the transcript or complex structure, length- or sequence-dependent changes may be observed in the cleavage of the transcript. We have done preliminary experiments on the NusA-modified *E. coli* RNA polymerase without observing strong effects. It is also possible that the unmodified polymerase•DNA•RNA ternary complex can exist in altered, stable conformational states, as suggested by the recent work of Telesnitsky and Chamberlin (1989) and Goliger *et al.* (1989) on factor-independent antitermination. Changes in the transcript structure at the target terminator might be detectable using a modified terminator including a self-cleaving structure. The system could easily be modified to include factor recognition sequences within the stems or loops of the “hammerhead” or upstream of the 5' end of the “hammerhead” structure. Coupled transcription and translation of a “hammerhead” could be used to study the interaction of

the ribosome with the nascent transcript. The method could be used for less well-defined transcription complexes such as eukaryotic RNA polymerase II or viral T7 RNA polymerase, and allowing for cross species comparisons. Part two of this chapter presents the results of studying T7 RNA polymerase using this method. A more applied use of the method is in the production of transcripts with defined 5' or 3' ends (Dzianott and Bujarski, 1988).

Our results are consistent with the generally accepted model of the transcription bubble, which holds that an RNA-DNA hybrid of 12 ± 2 nucleotides forms upstream of the active site for polymerization. The RNA in the hybrid would be unavailable for RNA-RNA structure formation, and therefore the transcript cannot self-cleave until the essential nucleotides of its final stem emerge from the hybrid. Our results actually require only that the RNA be restrained by interaction with another component of the ternary complex. Though these results do not directly demonstrate the hybrid's presence, in the remainder of the discussion we assume its existence. It is clear that the RNA is free to form structure immediately upon exiting the hybrid, suggesting that it does not form strong, specific interactions with the RNA polymerase outside this region. Photoaffinity results showing efficient labeling of RNA polymerase subunits by RNA's up to 94 nt in length indicate only that the RNA is spatially close to the protein; they do not indicate specific interactions (Bernhard and Meares, 1986).

Relevance to Termination

Models for transcription termination propose that an RNA-RNA hairpin is required for termination. Competition between RNA-RNA and RNA-DNA structure has been proposed to cause pausing and destabilization of the ternary complex. Our results suggest that the six-base final stem of the self-cleaving "hammerhead" structure cannot disrupt the RNA-DNA hybrid, that RNA-RNA and RNA-DNA structures of comparable stability do not interconvert rapidly. If such interconversion were facile, we would expect to see

cleavage starting when transcripts were elongated a minimum of 12 nucleotides beyond the cleavage site, when the nucleotides composing the final stem of the "hammerhead" reach the upstream end of the RNA-DNA hybrid. When less than 12 nucleotides have been synthesized beyond the cleavage point, disruption of the hybrid would require melting base pairs at the upstream end of the bubble for which no apparent Watson-Crick complement for the RNA is available, and it is reasonable that no cleavage would be observed even if there were equilibration. However, once 12 nucleotides have been made, formation of the final stem only requires that 6 base pairs of Watson-Crick RNA-RNA hybrid replace 6 base-pairs of RNA-DNA hybrid, and since these helices would have comparable stability the cleaving structure should form readily. These ideas are illustrated in Figure 6. The shortest transcript which cleaves completely, C₆₂, is shown in Figure 6(A) with a 13 nt RNA-DNA hybrid, and the shortest transcript which cleaves under any conditions within the complex, C₅₈, is shown in Figure 6(B) with a 12 nt hybrid. If the RNA-DNA hybrid were being disrupted during cleavage of transcript C₅₈, we would expect that the structure shown in Figure 6(C) could also form, and such a structure would give cleavage at position U₅₅, which is not observed. Our model is consistent with recent studies on pausing during transcription of *S. typhimurium* *his* operon mutants which suggest that the RNA-DNA hybrid is not completely disrupted at a strong pause site (Chan and Landick, 1989)

Since cleavage is only observed at lengths for which the hybrid can be essentially undisturbed, we propose that disruption of the hybrid may be prevented by the polymerase. These considerations and the stability of transcripts shorter than 57 nucleotides within the complex, even at 50 °C, suggest that the polymerase may specifically bind to and stabilize the RNA-DNA hybrid in the region before the proposed "separator" which disrupts the hybrid (Yager and von Hippel, 1987). We believe that the data on the temperature dependence of cleavage, showing approximately a 4 base range in which cleavage becomes more efficient at higher temperatures, reflect more rapid cleavage of molecules with truncated final stems at higher temperatures rather than more rapid equilibration between

Figure 6. Proposed Self-Cleaving Structures

The nascent RNA transcript, the DNA of the transcription bubble, and the RNA-DNA hybrid are shown. The cleavage point is indicated between nucleotides 43 and 44. The question mark represents the tertiary structure of the "hammerhead". The melted region of the DNA helix is shown as approximately 18 base pairs (Gamper and Hearst, 1982), and the active site for polymerization is shown at the front of the transcription bubble (Shi *et al.*, 1988); neither of these assumptions affect the interpretation of our results. The length of the RNA-DNA hybrid and the 3' nucleotide are indicated for each structure.

(A) The proposed structure of transcript C₆₂, the shortest transcript which cleaves completely at 22 °C.

(B) The structure of C₅₈, the shortest transcript which can cleave at temperatures up to 50 °C.

(C) A hypothetical structure, not believed to form, in which the RNA-RNA stem invades the RNA-DNA hybrid, yielding cleavage of the U₅₅ transcript. The total number of base-pairs is the same as would be found in a complex with a 12-base RNA-DNA hybrid.

Figure 6. Proposed Self-Cleaving Structures

complexes with complete final stems, because experiments on the RNA freed from the complex indicate that three nucleotides of the final stem are sufficient to allow cleavage (Fig. 4).

Of course, a terminator must eventually cause the disruption of the RNA-DNA hybrid. Our results imply that the hairpins in pause and termination sequences may need to have some minimum level of stability before they can overcome the binding energy of the polymerase for hybrid and thereby disrupt the hybrid. Perhaps hairpins of some minimum length are required in order to allow disruption, and there may be kinetic as well as thermodynamic barriers to termination depending on the specific hybrid base-pairs which must be disrupted during hairpin formation.

Part Two: Extension of the Transcript Folding Assay to T7 RNA Polymerase

Introduction:

In part two of this chapter we extend the cleaving transcript assay to the T7 RNA polymerase•DNA•RNA ternary complex during the elongation phase of transcription. The experiments address the question of when the RNA transcript in the T7 ternary elongation complex becomes free of steric interactions and available for structure formation with itself, allowing us to estimate the maximum possible length of the RNA–DNA hybrid in the ternary elongation complexes of T7 RNA polymerase. Comparison of the results for T7 and *E. coli* RNA polymerase will help determine whether the *E. coli* value of 12 nucleotides is a fundamental feature of RNA polymerase complexes.

T7 RNA polymerase is a small, monomeric bacteriophage enzyme with a molecular weight of about 100,000 that shows a high specificity for its promoters, which are well conserved (Chamberlin and Ryan, 1982). In contrast to the larger, multiple subunit bacterial and eucaryotic RNA polymerases, T7 RNA polymerase is much less complex both structurally and functionally and carries out all the steps of transcription by itself in vitro without the aid of additional protein factors. T7 RNA polymerase also shows homology to T3, SP6, and yeast mitochondrial RNA polymerases (Masters *et al.*, 1987). These factors combine to make T7 RNA polymerase an ideal system in which to study the “minimum apparatus” for transcription.

The maximum size of DNA protected by T7 RNA polymerase in the elongation mode has been determined by footprinting studies to be about 20 nucleotides. This establishes an upper limit to the size of the RNA-DNA hybrid. For *E. coli* RNA polymerase the size of the unwound region of the DNA in the ternary complex is about 17 nucleotides and the length of the RNA-DNA hybrid is about 12 nucleotides (Gamper and Hearst, 1982; Hanna and Meares, 1983; Monforte *et al.*, 1990). In view of the fact that T7 RNA polymerase is a smaller and simpler enzyme we expect that the maximum possible size of the RNA-DNA hybrid in its ternary complexes would be either the same or smaller than the *E. coli* RNA polymerase RNA-DNA hybrid length of 12 nucleotides. Our results demonstrate that for T7 RNA polymerase the RNA transcript is not free of interactions in the ternary complex until it is at least 10 ± 1 nucleotides away from the site of polymerization.

Materials and Methods:

(a) Materials

See part one.

(b) DNA Templates

A 127 nucleotide transcription template containing a Class III T7 promoter followed by the self-cleaving transcript (figure 1, in part 1) was synthesized from DNA oligomers 45 to 57 nucleotides in length by standard ligation procedures. This 127 nucleotide fragment containing the T7 promoter was next inserted into the polylinker region of puc19, which had been restricted with Bam HI and Eco RI and the resulting recombinant DNA was used to transform *E. coli* HB101. The desired plasmid was isolated by standard procedures and then digested with Eco RI to linearize the plasmid so as to produce a full-length run off transcript of 99 nucleotides, on transcription.

(c) In vitro Transcription Reactions

All transcripts were initiated using [α - 32 P]pGpG so that the 5' end of the RNA was labelled. The dinucleotide GpG was kinased according to the method in part one. The kinase was heat killed for 3 minutes at 65 °C and the buffer conditions were adjusted to those appropriate for transcription by T7 RNA polymerase by the addition of a conversion buffer. This mixture was further diluted by the addition of DNA template, RNA polymerase, cold GpG (final concentration 180 μ M) and this mixture was then aliquoted to the individual reaction tubes.

Transcription reactions were performed at 37 °C for 20 minutes in 5 ml volumes in T7 transcription buffers: 40 mM TrisOAc, pH 8.0, 100 mM NaCl, 10 mM spermidine, 0.1% Triton X, 100 mM DTT, and 15 mM MgCl₂. T7 RNA polymerase was at a concentration of 50 nM and the DNA template was in 50% excess, being 75 nM per lane.

The concentration of NTPs ranged from 1-50 mM (actual concentrations in figure legends). The 3' dNTPs were in the range of 5-20 mM.

T7 RNA polymerase and DNA were added to [α - ^{32}P]pGpG mixture containing cold GpG, RNasin and the appropriate buffer. This was incubated for 8 minutes and then aliquoted into eppendorf tubes containing various nucleoside triphosphates and chain terminator mixtures. The reaction mixtures were then incubated for 20 minutes and the transcription reaction was terminated by the addition of equal volume of formamide containing 20 mM EDTA and dyes. These were then analyzed by electrophoresis on 12% polyacrylamide gels as in part one.

In experiments where the ternary complex was disrupted and the released transcripts self-cleave, reaction mixture was boiled for 1.5 minutes, rapidly cooled to room temperature and incubated for 30 minutes, the formamide-EDTA solution was added and the mixture was analyzed.

Results:

Transcription reactions were performed on DNA templates with a strong T7 ClassIII promoter. Only RNA molecules which had been initiated with [α - ^{32}P]pGpG were detected. Since the label is at the 5' end of the transcript, only the 5' cleavage products of the RNA molecules are visible in these experiments. Transcriptions were performed in the presence of chain terminating nucleotides, the 3'-deoxyribonucleoside triphosphates. A sequencing ladder of stalled complexes was created by using the four nucleotide analogs in separate reactions. Transcription reactions were quenched with a mixture of formamide and EDTA in order to simultaneously disrupt the ternary complex and prevent further "hammerhead" cleavage by denaturing the RNA and sequestering all divalent cations.

Two different templates were used, one coding for the "hammerhead" sequence and one control containing a single base change within the "hammerhead" consensus sequence (Fig. 1, in part 1). The control template produces a transcript which does not cleave. The templates used were identical to the ones used in part one except that a T7 promoter has replaced the *E. coli* Tac 18 promoter. The "hammerhead" and control T7 templates were cloned into the polylinker site of puc19 and the modified plasmid was then restricted with Eco RI so that upon transcription it produced run-off transcription products.

The same basic requirement for ternary complex stability during the course of the experiment needed for the *E. Coli* RNA polymerase study apply for the T7 RNA polymerase ternary complex. The ternary complex must be stable for a longer period of time than the complex takes to transcribe, stall, and form RNA structure. Elongation by one base takes 4.3 msec for T7 RNA polymerase (Golomb and Chamberlain, 1974). The rate of RNA structure formation is comparable and is also in the millisecond range (Puglisi, 1989). The half-life of the isolated hammerhead is less than 30 seconds (data not shown). T7 RNA polymerase ternary complexes have not been well studied but the complex is stable for the course of an hour (S. Sastry, personal communication) and the heat/cool experiment presented below confirms that the ternary complex is stable for the course of the

experiment. During the course of a 30 minute experiment the ternary complex has ample time to transcribe, stall and allow RNA structure to form, providing a clear picture of extent of cleavage at each position and the influence of the ternary complex on cleavage.

Transcript self cleavage in stalled elongation complexes:

Figure 7 shows a typical autoradiogram of sequencing reactions performed on the "hammerhead" and non-cleaving control templates using T7 RNA polymerase. The control template produces a standard sequencing pattern of stalled complexes for the full length of the DNA template (Lanes 1-5). The pattern of stalled complex transcripts from the "hammerhead" (Lanes 6-10) is identical to the pattern from the control template (Lanes 1-5) for the first 42 nucleotides after which the pattern changes due to transcript self-cleavage. The strong band at C₄₃ (Lanes 6-15) is the cleavage product (C₄₃ indicates the transcript of length 43, whose last base is a C). From A₄₄ to U₅₅ the pattern is similar again to the control template indicating that transcripts from the "hammerhead" template are not efficiently cleaved in this range. Beyond U₅₅ the band intensity is dramatically reduced and is almost negligible indicating there is nearly 100% cleavage in complexes of longer transcript lengths. Certain bands between G₄₃-C₅₆ have diminished intensity especially C₄₉. The diminished intensity of these bands indicate a greater level for cleavage in these transcript lengths. This may be due to the particular sequence of the RNA in these ternary complexes which makes them unusually unstable and causes the complex to fall apart releasing free RNA, which then self-cleaves.

Figure 7. Transcript Self-Cleavage in the T7 Elongation Complex

Comparison of RNA sequencing ladders from transcription of the control and "hammerhead" templates show self-cleavage of the transcript. Lanes 1-5 show transcription on the control template in the presence of four different chain terminators. Lanes 6-10 are equivalent to 1-5 except the "hammerhead" template coding for the self-cleaving transcript is used. Lanes 11-15 are identical to 6-10 except that it shows transcript self-cleavage after disruption of ternary complex.

Transcription was performed at 37°C for 30 minutes. The triphosphate concentrations used in each lane are as follows:

Control and "hammerhead" experiments (lanes 1-15). Control lanes: ATP, UTP, and CTP 400 μ M; GTP 50 μ M. U lane: 3'-dATP 100 μ M; ATP 400 μ M; GTP 50 μ M; CTP and UTP 1 mM. U lanes: 3'-dUTP 50 μ M; UTP 400 μ M; GTP 50 μ M; CTP and ATP 1 mM. G lanes: 3'-dGTP 5 μ M; GTP 100 μ M; UTP, ATP, and CTP 1 mM. C lanes: 3'-dCTP 50 μ M; CTP 400 μ M; GTP 50 μ M; UTP and ATP 1 mM.

In Lanes 6-15 transcription was performed on "hammerhead" template at 37 °C for 20 minutes at which time the reaction volume were split. For lanes 6-10 the reactions were immediately quenched with EDTA/formamide. For lanes 11-15 the reactions were heated to 90 °C for 2 minutes to disrupt the complex and release the transcripts, quick cooled to 22 °C, incubated for 30 minutes, and then quenched. This procedure allows free transcripts of sufficient length to self-cleave.

Transcript Self-Cleavage in the T7 Elongation Complex

Figure 7

The heat/cool experiment of lanes (11-15) demonstrate that the integrity of the ternary complex is essential for prevention of the self-cleavage. When the ternary complex is disrupted by heating, the RNA transcript is released and the bands above G₄₆ disappear. Thus molecules long enough to form 3 or more base pairs of the final "hammerhead" stem self-cleave. This demonstrates that the ternary complex is stable over the course of the experiment because released RNA molecules would rapidly self-cleave. The presence of ternary complex is essential for non-cleavage of transcripts.

The cut-off position is reproducible and is obtained for a wide range of concentration of NTPs and dNTPs. It is not altered whether a supercoiled or linearized plasmid is used. Temperature studies indicated that the same cut-off is obtained at 30 °C and 37 °C. At 23 °C several intermediates of lengths between 56 and 99 nucleotides are observable (data not shown). This could be attributed to a slower rate of cleavage, due to the stabilization of alternate conformations, at this temperature. Time course studies of the cleavage reaction show that the ternary complex is stable throughout the 20 minute range and the same cut-off is obtained at different time points (data not shown).

A minimum length of the first 46 nucleotides of the RNA transcript is essential for cleavage. These results demonstrate that the formation of "hammerhead" structure and self-cleavage of transcripts is inhibited for RNA molecules within the ternary complex up to 55 nucleotides in length. Self-cleavage of the transcripts occur only when 10 nucleotides have been synthesized beyond G₄₆. Thus when the transcript is elongated to C₅₆, the "hammerhead" sequence becomes capable of forming the final stem of the "hammerhead" structure, allowing cleavage to occur. The separation between the site of polymerization and G₄₆ is at least 10 nucleotides, this means that a maximum of 10 nucleotides of the transcript are unavailable for RNA-RNA base pair formation, beyond which the transcript is free to form structure.

Discussion:

Our data clearly demonstrate that in the T7 RNA polymerase•DNA•RNA ternary complex the formation of intramolecular structure in the nascent RNA transcript is inhibited for the first 10 ± 1 nucleotides upstream of the site of polymerization, whereas beyond 10 nucleotides RNA structure can form freely. This is distinctly different from the 12 nucleotide RNA–DNA hybrid length currently accepted in the literature for *E. coli* RNA polymerase and which was also confirmed using this same assay, implying that a 12 nucleotide RNA–DNA hybrid is not a fundamental feature of the transcription apparatus and may vary depending on the size and complexity of the RNA polymerase. It also introduces a new minimum to the possible length of the RNA–DNA hybrid and suggests that an elongating ternary complex can be stabilized by a shorter RNA–DNA hybrid of 10 nucleotides. The data fit in with the picture of T7 RNA polymerase being a "minimum apparatus" for transcription.

Extension of these studies to the more complex eucaryotic RNA polymerases would prove informative. Recent RNase digest experiments have indicated that 18-22 nucleotides are protected from RNase digestion by RNA Pol II (Choder and Aloni, 1988). This result suggests that either the length of RNA protected due to interactions in the ternary complex is much larger than *E. coli* or T7 RNA polymerase or that large exogenous probes like RNase is sterically excluded from the site of transcript emergence. While such an experiment can provide an outer limit to the length of protected RNA, it cannot clearly identify the point at which the RNA is able to interact with itself and where polymerase specific, transcript binding accessory factors will have access to the nascent transcript.

Footprinting studies by Ikeda and Richardson (1986) have shown that the conformation of T7 RNA polymerase is different after 15 nucleotides of synthesis than it is after 6 nucleotides. These data suggest that a switch between initiation and elongation modes must occur somewhere in this range. Studies of abortive initiation by T7 RNA polymerase (Martin *et al.*, 1988) show that the maximum length of abortive transcription is 8

nucleotides, suggesting that an abrupt change in the stability of the ternary complex occurs upon the incorporation of 9th nucleotide. It is likely that the transition to a stable elongation complex occurs at 9 or at most 10 nucleotides into transcription, since beyond 10 nucleotides the RNA exits from the transcription complex.

Studies on isolated stalled complexes of *E. coli* RNA polymerase have shown that when the RNA transcript length is increased from 8 to 10 nucleotides the RNA polymerase exhibits a change in conformation and the loss of sigma factor, this corresponds to a shift from the initiation complex to the more stable elongation mode (Krummel and Chamberlin, 1989). The stabilization of the ternary complex occurs when the transcript length is about 9-10 nucleotides, even though the length of the RNA-DNA hybrid in the elongation mode is 12 nucleotides. This suggests that an RNA-DNA hybrid length of 10 nucleotides is sufficiently long to stabilize the ternary elongation complex of *E. coli* RNA polymerase.

The length of 10 nucleotides is significant in the stabilization of the ternary elongation complexes of both T7 RNA polymerase and *E. coli* RNA polymerase. This length represents the minimum and sufficient length of RNA needed for stabilization of the RNA polymerase ternary complexes in the stable elongation mode and may represent the transition point from initiation to elongation for all RNA polymerases. Extension of the length of the RNA-DNA hybrid by polymerases such as *E. coli* RNA polymerase increases the stability of the complex and allows for more refined control over different aspects of transcription regulation, especially termination. Such fine tuning is not required by simple viral systems such as T7.

References:

- Berkhout, B., Silverman, R. H., and Jeang, K.-T. (1989). "Tat *Trans*-Activates the Human Immunodeficiency Virus through a Nascent RNA target." *Cell* **59**, 273-282.
- Bernhard, S. L. and Meares, C. F. (1986). "Accessibility of the Leading End of Ribonucleic Acid in Transcription Complexes." *Biochemistry* **25**(21), 6397-6404.
- Brennan, C. A., Dombroski, A. J., and Platt, T. (1987). "Transcription Termination Factor Rho Is an RNA-DNA Helicase." *Cell* **48**, 945-952.
- Buzayan, J. M., Feldstein, P. A., Segrelles, C., and Bruening, G. (1988). "Autolytic Processing of a Phosphorothioate Diester Bond." *Nucl. Acids Res.* **16**(9), 4009-4023.
- Chamberlin, M. and Ryan, T. (1982). "Bacteriophage RNA Polymerases." *The Enzymes* **15**, 87-108.
- Chamberlin, M. J. (1974). "Bacterial DNA-Dependent RNA Polymerase." In *The Enzymes*. Ed. Boyer, P. D. New York: Academic Press. Vol. 10, pp. 333-374.
- Chan, C. L. and Landick, R. (1989). "The *Salmonella typhimurium* *his* Operon Leader Region Contains an RNA Hairpin-dependent Transcription Pause Site: Mechanistic implications of the effect on pausing of altered RNA hairpins." *J. Biol. Chem.* **264**(34), 20796-20804.
- Choder, M. and Aloni, Y. (1988). "RNA Polymerase II Allows Unwinding and Rewinding of the DNA and thus maintains a constant length of the Transcription Bubble." *J. Biol. Chem.* **263**(26), 12994-13002.
- Dzianott, A. M. and Bujarski, J. J. (1988). "An *in Vitro* Transcription Vector which Generates Nearly Correctly Ended RNAs by Self-cleavage of Longer Transcripts." *Nucl. Acids Res.* **16**(22), 10940.
- Eperon, L. P., Graham, I. R., Griffiths, A. D., and Eperon, I. C. (1988). "Effects of RNA Secondary Structure on Alternative Splicing of Pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase?" *Cell* **54**, 393-401.
- Ericson, G. and Wollenzien, P. (1989). "An RNA Secondary Structure Switch between the Active and Inactive Conformations of the *Escherichia coli* 30S Ribosomal Subunit." *J. Biol. Chem.* **264**(1), 540-545.
- Farnham, P. J. and Platt, T. (1980). "A Model for Transcription Termination Suggested by Studies on the *trp* Attenuator *In Vitro* Using Base Analogs." *Cell* **20**, 739-748.
- Faus, I., Chen, C.-Y. A., and Richardson, J. P. (1988). "Sequences in the 5' Proximal Segment of the Paused Transcript Affect NusA-mediated Enhancement of Transcriptional Pausing." *J. Biol. Chem.* **263**(22), 10830-10835.

- Fedor, M. J. and Uhlenbeck, O. C. (1990). "Substrate Sequence Effects on "Hammerhead" RNA Catalytic Efficiency." *Proc. Natl. Acad. Sci. USA* **87**, 1668-1672.
- Forster, A. C. and Symons, R. H. (1987). "Self-Cleavage of Plus and Minus RNAs of a Virusoid and a Structural Model for the Active Sites." *Cell* **49**, 211-220.
- Gamper, H. B. and Hearst, J. E. (1982). "A Topological Model for Transcription Based on Unwinding Angle Analysis of *E. coli* RNA Polymerase Binary, Initiation and Ternary Complexes." *Cell* **29**, 81-90.
- Golomb, M. and Chamberlin, M. (1974). *Biochemistry* , 2858-2863.
- Hanna, M. M., Dissinger, S., Williams, B. D., and Colston, J. E. (1989). "Synthesis and Characterization of 5-[(4-Azidophenacyl)thio]uridine 5'-Triphosphate, a Cleavable Photo-Cross-Linking Nucleotide Analogue." *Biochemistry* **28**(14), 5814-5820.
- Hanna, M. M. and Meares, C. F. (1983). "Topography of Transcription: Path of the leading end of nascent RNA through the *Escherichia coli* transcription complex." *Proc. Nat. Acad. Sci. USA* **80**, 4238-4242.
- Ikeda, R. A. and Richardson, C. C. (1986). "Interactions of the RNA Polymerase of Bacteriophage T7 with its Promoter during Binding and Initiation of Transcription." *Proc. Natl. Acad. Sci. USA* **83**, 3614-3618.
- Kahn, J. D. and Hearst, J. E. (1989). "Reversibility of Nucleotide Incorporation by *Escherichia coli* RNA Polymerase, and Its Effect on Fidelity." *J. Mol. Biol.* **205**, 291-314.
- Keese, P. and Symons, R. H. (1987). "The Structure of Viroids and Virusoids." In Viroids and Viroid-Like Pathogens. Ed. Semancik, J. S. Boca Raton, Florida: CRC Press. pp. 1-47.
- Kramer, F. R. and Mills, D. R. (1981). "Secondary Structure Formation during RNA Synthesis." *Nucl. Acids Res.* **9**(19), 5109-5124.
- Krummel, B. and Chamberlin, M. J. (1989). "RNA Chain Initiation by *Escherichia coli* RNA Polymerase. Structural transitions of the enzyme in early ternary complexes." *Biochemistry* **28**(19), 7829-7842.
- Kumar, S. A. and Krakow, J. J. (1975). "Studies on the Product Binding Site of the *Azotobacter vinelandii* Ribonucleic Acid Polymerase." *J. Biol. Chem.* **250**(8), 2878-2884.
- Landick, R. and Yanofsky, C. (1987). "Isolation and Structural Analysis of the *Escherichia coli trp* Leader Paused Transcription Complex." *J. Mol. Biol.* **196**, 363-377.

- Levin, J. R., Krummel, B., and Chamberlin, M. J. (1987). "Isolation and Properties of Transcribing Ternary Complexes of *Escherichia coli* RNA Polymerase Positioned at a Single Template Base." *J. Mol. Biol.* **196**, 85-100.
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Martin, C. T., Muller, D. K., and Coleman, J. E. (1988). "Processivity in Early Stages of Transcription by T7 RNA Polymerase." *Biochemistry* **27**(11), 3966-3974.
- Masters, S. B., Strohl, T. L., and Clayton, A. D. (1987). "Yeast mitochondrial RNA polymerase is homologous to those encoded by T3 & T7." *Cell* **51**, 89-99.
- Monforte, A. J., Kahn, D. J., and Hearst, E. J. (1990). "RNA folding during transcription by *Escherichia coli* RNA polymerase Analyzed by RNA Self-cleavage." *Biochemistry* **29**, 7882-7890.
- Mulligan, M. E., Brosius, J., and McClure, W. R. (1985). "Characterization *in Vitro* of the Effect of Spacer length on the Activity of *Escherichia coli* RNA Polymerase at the TAC Promoter." *J. Biol. Chem.* **260**, 3529-3538.
- Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R., and Bruening, G. (1986). "Autolytic Processing of Dimeric Plant Virus Satellite RNA." *Science* **231**, 1577-1580.
- Puglisi, J. D. (1989). RNA Folding: Structure and Conformational Equilibria of RNA Pseudoknots. Dissertation, University of California, Berkeley.
- Rothman, J. E. (1989). "Polypeptide Chain Binding Proteins: Catalysts of protein folding and related processes in cells." *Cell* **59**, 591-601.
- Shi, Y.-b., Gamper, H., Van Houten, B., and Hearst, J. E. (1988). "Interaction of *Escherichia coli* RNA Polymerase with DNA in an Elongation Complex Arrested at a Specific Psoralen Crosslink Site." *J. Mol. Biol.* **199**, 277-293.
- Stackhouse, T. M. and Meares, C. F. (1988). "Photoaffinity Labeling of *Escherichia coli* RNA Polymerase/Poly[d(A-T)] Transcription Complexes by Nascent RNA." *Biochemistry* **27**(8), 3038-3045.
- Uhlenbeck, O. C. (1987). "A Small Catalytic Oligoribonucleotide." *Nature* **328**, 596-600.
- von Hippel, P. H., Bear, D. G., Morgan, W. D., and McSwiggen, J. A. (1984). "Protein-Nucleic Acid Interactions in Transcription: A Molecular Analysis." *Annu. Rev. Biochem.* **53**, 389-446.
- Williams, A. L., Jr. and Tinoco, I., Jr. (1986). "A Dynamic Programming Algorithm for Finding Alternative RNA Secondary Structures." *Nucl. Acids Res.* **14**(1), 299-315.

- Yager, T. D. and von Hippel, P. H. (1987). "Transcript Elongation and Termination in *Escherichia coli*." In *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. Ed. Neidhardt, F. C. Washington, D.C.: Am. Soc. Microbiol. pp. 1241-1275.
- Yager, T. D. and von Hippel, P. H. (1989). "A Thermodynamic Model of Transcription Termination." *Biochemistry* , in press.
- Yang, X. and Roberts, J. W. (1989). "Gene *Q* Antiterminator Proteins of *Escherichia coli* Phages 82 and λ Suppress Pausing by RNA Polymerase at a p-Dependent Terminator and at Other Sites." *Proc. Nat. Acad. Sci. USA* 86, 5301-5305.

END

**DATE
FILMED
9 / 3 / 92**