Title

REACTION OF PSORALEN WITH RNA: SPECIFICITY AND USE AS A PROBE FOR SECONDARY STRUCTURE ANALYSIS

Author

Thompson, J.F.

Publication Date

1982-09-01

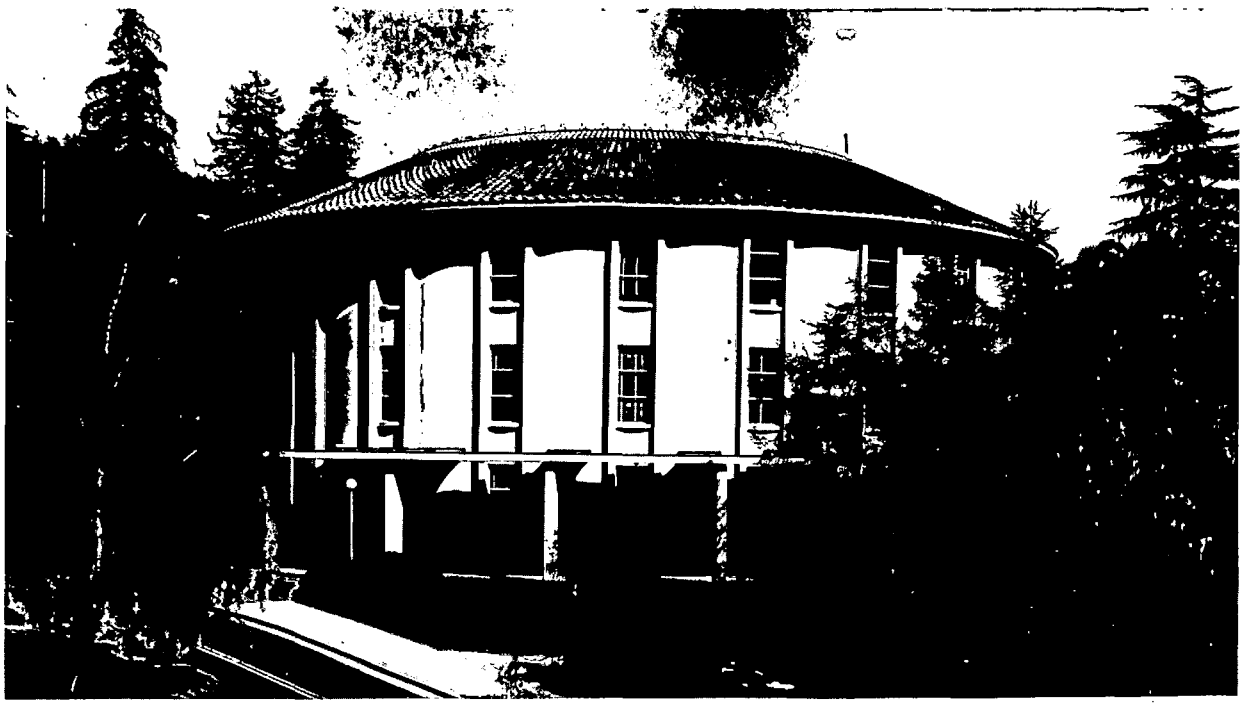
Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

CHEMICAL BIODYNAMICS DIVISION

REACTION OF PSORALEN WITH RNA: SPECIFICITY AND USE AS A PROBE FOR SECONDARY STRUCTURE ANALYSIS

John Fenton Thompson
(Ph.D. thesis)

September 1982



Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

LBL-14991

DISCLAIMER

LBL-14991

REACTION OF PSORALEN WITH RNA: SPECIFICITY AND USE AS A PROBE FOR SECONDARY STRUCTURE ANALYSIS*

John Fenton Thompson
(Ph.D. thesis)

Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

September 1982

*This work was supported by the U. S. Department of Energy under Contract No. DE-AC03-76SF00098.
This manuscript was printed from originals provided by the author.

John Fenton Thompson

Reaction of Psoralen with RNA: Specificity and Use as a Probe for Secondary Structure Analysis

ABSTRACT

A variety of techniques have been used to study how psoralen and its derivatives react with RNA. This information has then been used to analyze the secondary structure of different ribosomal RNAs. The structural information obtained has allowed us to make several proposals concerning the nature of ribosomal functions.

Paper electrophoresis at pH 3.5 and 8.8 and HPLC have been used to get high resolution separation of RNA-psoralen adducts. The separated adducts have been analyzed and shown to be primarily uridine adducts with the psoralen reacted at the furan end. The stereochemistry of the major adducts was determined by NMR. The effect of structural transitions on the number and type of adducts was found for several polymers. Low salt and temperature, in general, produced maximal incorporation. The effect of psoralen structure on crosslinking ability was analyzed. Charged derivatives formed monoadducts very efficiently but did not produce the level of crosslinking obtainable with lower levels of reaction with uncharged derivatives.

Secondary structure analysis of D. melanogaster 5S RNA yielded two definite and two tentative crosslinks which support the generally accepted models for 5S structure. Analysis of E. coli 16S RNA by gel techniques yielded 13 cross-

links. Of these, seven are present in all published secondary structure models, three are present in one or two, and three are new. The positions of these crosslinks have allowed us to make detailed structure-function correlations for tRNA proofreading and translocation. Evidence is also presented for an interaction between eukaryotic mRNA (5' cap structure) and 18S RNA (hypermodified base amΨ) which serves a function analogous to the Shine-Dalgarno sequence in prokaryotes.

John E. Hearst

DEDICATION

To Patty

TABLE OF CONTENTS

Acknowledgements

This thesis would not have been possible without the continual help and support of many people throughout the last four years. I cannot hope to thank all those who made my stay at Berkeley so stimulating, but this thesis is proof that their assistance did not go for nought.

I am especially grateful to my original office-mates; Corliss Chun, Rick Reisbig, and Doug Youvan; who instilled in me a more critical and realistic view of science. I learned much from their endless arguments on all types of problems. Other scientists with whom I have worked closely have also imprinted their attitudes and knowledge on me. My work has benefitted greatly from working with Steve Isaacs, Maurice Wegnez, George Cimino, and Jean-Pierre Bachellerie.

The comments, criticisms, and friendship of others in the various parts of the Hearst lab were also indispensable: Kris Zsebo, Gary Wiesehahn, Mark Walters, John Tessman, Sohel Talib, Diane Sandlin, Jane Rupple, Dave Kanne, Brian Johnston, Aku Ikoku, John Hubbard, Chiao-Chain Huang, Bruce Greenberg, Jeremy "Gold", Howard Gamper, J. T. Elder, Ed Bylina, Dean Bender, and Barry Ballard. The people who contributed technical assistance that made my work so much easier include Marion Malone, Kathy Macbride, Lisa Kona, Margaret Knight, Karen Elliot, and Steven Quong. Marshall, of course, goes in a separate category.

The Tinoco lab was a special asset. Not only did they help with many melting curves and CD spectra but they also

staged the infamous cold noodle parties and ensured that the Hearst lab was not overproductive. Special thanks to Steve Winkle, Milan Tomic (honorary), John Platenak, Art Pardi, Jeff Nelson, Kathy Morden, Gloria Meng, Helen Lok, David Koh David Keller, Joe Kao, Kathi Hall, Barbara Dengler, and Ken Dahl.

Most of all, this thesis is a tribute to John Hearst. His courage to let me decide my own fate in science and chase wild ideas while always keeping me on track has taught me more than constant supervision ever could have. I just hope I can always maintain the same enthusiasm for science that he has. Thanks John.

ABBREVIATIONS

A	adenine
acp^3U	3-(3-amino)-carboxypropyluridine
AMP	adenosine-3'-monophosphate
AMT	4'-aminomethyl-4,5',8-trimethylpsoralen
amψ	3-(3-amino)-carboxypropyl-1-methylpseudouridine
ATP	adenine-5'-triphosphate
BA	base adduct (p. 44)
BAP	bacterial alkaline phosphatase
Bis-MSB	p-bis(O-methylstyryl)benzene
BND	benzoylylatednapthylateddietylyaminoethyl
C	cytidine
CD	circular dichroism
CMP	cytidine-3'-monophosphate
CMT	4'-chloromethyl-4,5',8-trimethylpsoralen
cUMP	uridine-2',3'-cyclic phosphate
D1, D2	diastereomer one, diastereomer two
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
E1, E2	equilibrium product one, equilibrium product two
EDTA	ethylenediaminetetraacetic acid
EF-G	elongation factor G
EF-Tu	elongation factor Tu
G	guanine
GMP	guanine-3'-monophosphate
GDP	guanine-5'-diphosphate
GTP	guanine-5'-triphosphate

HMT	4'-hydroxymethyl-4,5',8-trimethylpsoralen
HPLC	high pressure liquid chromatography
IMT	4'-iodomethyl-4,5',8-trimethylpsoralen
LES	0.1 M LiCl, 10 mM EDTA, 0.5% SDS, 10 mM Tris base
m^7G	7-methylguanine
8-MOP	8-methoxypsoralen
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
OD	optical density at 260 nm
PMT	N-(4'-methylenetrioxsalen)-pyridinium chloride
PMX	N-(5-methylene-8-methoxypsoralen)-pyridinium chloride
poly(A)	polyadenylic acid
poly(A,U)	polyadenylicuridylic acid (random)
poly(A-U)	polyadenylicuridylic acid (strictly alternating)
poly(C)	polycytidylic acid
poly(G)	polyguanodylic acid
poly(I)	polyinosoic acid
poly(U)	polyuridylic acid
POPOP	1,4-bis[2-(5-phenyloxazolyl)]benzene
PPO	2,5-diphenyloxazole
R_F	relative mobility in paper electrophoresis at pH 3.5
R_F^*	relative mobility in paper electrophoresis at pH 8.8
SDS	sodium dodecyl sulfate
TEMED	N,N,N',N',-tetramethylethylenediamine

TMA I	10 mM $MgCl_2$, 100 mM NH_4Cl, 10 mM Tris-HCl (pH 7.2) 14 mM β-mercaptoethanol
TMP	4,5',8-trimethylpsoralen
Tris	tris(hydroxymethyl)aminomethane
U	uridine
UMP	uridine-3'-monophosphate
UTP	uridine-5'-triphosphate
UV	ultraviolet
VRC	vadanylribonucleoside complex
Ψ	pseudouridine

Chapter One:
Introduction

The unambiguous determination of RNA secondary structure is a problem that has plagued researchers for many years. The problem lies not in formulating reasonable structures, but in choosing from among the many available. The techniques most commonly used to clarify this situation, chemical modification and partial enzymatic digestion, have only been mildly successful in limiting the number of acceptable structures. These methods indicate which regions are most exposed to attack by enzymes or chemicals and thus must be single-stranded. Unlike other enzymes, cobra venom ribonuclease cuts specifically in double stranded regions, thus yielding additional information. None of these probes, however, gives direct information about which regions are paired to which. To get around this, two variations of partial enzymatic digestion have been employed. Rather than simply map the location of cutting sites, entire fragments have been isolated, denatured, and their components sequenced (Ross and Brimacombe, 1979; Glotz and Brimacombe, 1980; Glotz et al., 1981). This gives information about which regions are interacting in distant parts of the RNA. Care must be taken in interpreting this data because any time the RNA chain is cut, there is a possibility that rearrangements in structure will occur. The other variation of the enzymatic digestion method is to examine which RNA fragments are

protected when a protein which binds specifically to a particular RNA is attached. Sometimes, only short stretches of RNA are protected, but occasionally distant regions in a molecule are protected by a single protein (Rinke et al., 1976). As long as it can be shown that the protein can still specifically bind the digested fragment, it can be safely assumed that the interactions actually occur in vivo. This procedure is limited by the number of proteins which can be used.

More recently, phylogenetic comparisons have proved valuable. When used in conjunction with the aforementioned techniques, widely accepted models for the secondary structure of various ribosomal RNAs have been generated (Fox and Woese, 1975; Noller and Woese, 1981). This technique is based on the assumption that RNAs from different species which carry out the same function must have approximately the same secondary and tertiary structure despite differences in sequence. Therefore, it must be possible to generalize any proposed structure to all species for which the RNA sequence is known. While this method has been more powerful than any other in analyzing such structures as 16S RNA (Noller and Woese, 1981) and 23S RNA (Noller et al., 1981), it also has severe limitations. It is difficult to completely analyze very large RNAs and impossible to analyze RNA of which there is only one example available (viral RNAs, some mRNAs, etc.).

The only technique which can yield simple, direct in-

formation about the structure of any RNA is crosslinking. While many crosslinking reagents have the potential to be used as probes to RNA structure, only a few have been employed with any degree of success. Ultraviolet light has been used to determine a number of crosslinks to high resolution in 16S RNA (Zweib and Brimacombe, 1980). This information has proved very useful, but UV light has two disadvantages. It can cause extensive damage to both nucleic acids and proteins, and it is also not clear what the exact structural requirements are for crosslinking. Certainly, the crosslinked bases must be physically very close because there is no chemical reagent to bridge any distance. However, it is not known what kinds of secondary or tertiary interactions promote reaction.

Phenyl diglyoxal has also been used as a chemical crosslinker (Wagner and Garrett, 1978; Hancock and Wagner, 1982). While glyoxal generally only reacts with single-stranded guanines, the phenyl ring evidently allows intercalation of the compound because initial results yielded a crosslink between two obviously double-stranded guanines (Wagner and Garrett, 1978). Further results have only complicated the interpretation (Hancock and Wagner, 1982; see chapter 5 for further discussion). Most of the reactive sites observed thus far normally are unreactive to glyoxal so some unfolding of the RNA must be occurring. This ambiguity clearly limits the compound's usefulness. More recently, a dialdehyde connected by an aliphatic chain has been synthesized

to overcome this problem (H. Noller, personal communication). The crosslinks to be found, if any, will be between single-stranded regions. Crosslinked guanines will be close spatially, but adjacent bases might not be involved in any secondary or tertiary interactions.

Use of psoralen as a crosslinker circumvents all of the problems mentioned above. Crosslinking is done in combination with long-wave ultraviolet light and thus causes no damage to the remainder of the RNA. Psoralen is reacted with the intact RNA in whatever conditions are necessary so native structure can be maintained. Crosslinks are limited to double-stranded regions, and the reaction between nucleic acids and psoralen has been well studied.

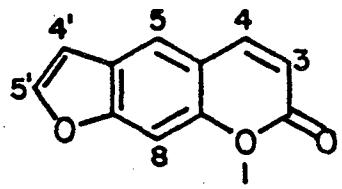
Psoralen and its derivatives have been used for over 3000 years to treat skin diseases such as psoriasis (reviewed by Scott et al., 1976). Its use as a tool for nucleic acid structure analysis began somewhat more recently when it was discovered that the basis for psoralen's action arises from its ability to covalently crosslink DNA via the 5, 6 double bond of pyrimidines on opposite strands of the helix (Dall' Acqua et al., 1970; Cole, 1971). The sequence of reactions which lead to crosslink formation are now fairly well characterized.

The first reaction to occur is the non-covalent intercalation of the psoralen between base pairs of DNA or RNA. This intercalation, or dark binding, is accompanied by an unwinding of the helix with a concomitant lengthening of the

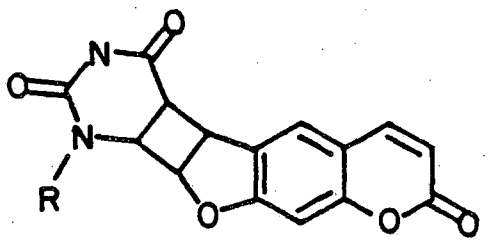
of the polymer. The amount of unwinding is comparable to other known intercalators (Wiesehahn and Hearst, 1978). Flow dichroism experiments have confirmed the intercalative mode of dark binding (Dall' Acqua et al., 1978; Tjerneld et al., 1979). The rate constant for dark binding is highly dependent on a number of factors. Effects of substituents at various positions on the psoralen heterocycle have been examined (Isaacs et al., 1977, 1982). Positively charged derivatives bind much more strongly than uncharged derivatives because, in addition to hydrophobic and base stacking interactions, there is also a strong electrostatic attraction. Conditions which stabilize the helix, such as high salt, lower the rate of intercalation. The geometry of the dark bound psoralen places severe constraints on how it may react. Even before definitive structural analyses were performed on psoralen adducts, it was clear what stereoisomers should be allowed.

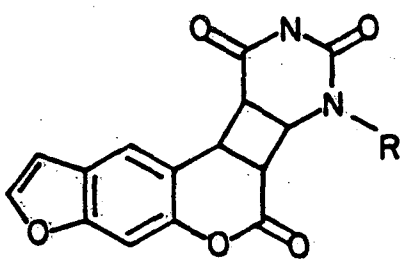
The photoreactive bonds of psoralen are the 4', 5' and 3, 4 double bonds (see figure 1 for structure and numbering system). After absorption of 300-400 nm light, either of these can photo-react and form a cyclobutane bridge to the 5, 6 double bond of a pyrimidine base. Purines can also react, but to a much lesser extent (chapter 2). Reaction with proteins has been reported, but no protein adducts have ever been isolated (Pathak, et al., 1974). If reaction actually does occur, it is at a much lower rate than with nucleic acids.

Psoralen



Monoadducts





Crosslink

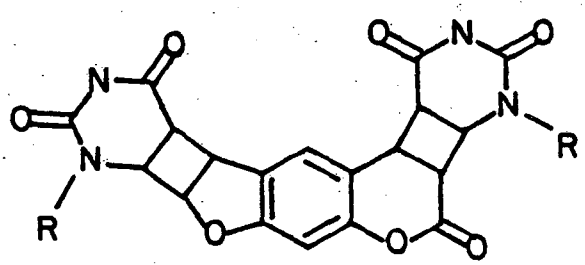


Figure 1: The structure and numbering system for psoralen is shown at the top. The two possible monoadducts are diagrammed in the middle. The stereochemistry around the cyclobutane ring is discussed in chapter 2. A crosslink between two uridines is shown at the bottom.

Theoretical calculations predicted that the 3, 4 double bond would be the most reactive (Song et al., 1971). These did not take the effect of the DNA helix into account, however. It is now clear that the 4', 5' bond is actually more reactive in the DNA (Straub et al., 1981). Unlike the 3, 4 adduct, the 4', 5' adduct retains significant absorbance above 300 nm. It can thus absorb a second photon to form a crosslink providing there is a pyrimidine suitably located on the opposite strand of the helix. Crosslink formation was shown to be a two photon event by pulsed laser studies (Johnston et al., 1977). There is a relaxation time required, presumably needed for a conformational change, before most monoadducts can go on to form crosslinks (Johnston et al., 1981).

The relative reactivities of pyrimidines as polymers and as nucleosides in a variety of conditions have been examined. In all cases, thymine and uracil have been more reactive than cytosine (Krauch et al., 1967; Pathak et al., 1974; Bachellerie et al., 1981; Straub et al., 1981). The difference in reactivity is accentuated when the nucleosides are base paired in polymers. Reaction of nucleosides in solution or in frozen glasses yields a much larger number of stereoisomeric products than the reaction of polymers in solution (Straub et al., 1981).

Detailed structural analyses have recently been performed on psoralen adducts. High resolution separation of different stereoisomers has been made possible by HPLC

(Straub et al., 1981; Kanne et al., 1982a, b; chapter 2). NMR studies have shown that the stereochemistry of both monoadducts and crosslinks is as previously supposed. Mass spectroscopy, circular dichroism, ultraviolet absorption, and fluorescence studies have also been done on isolated adducts. Two separate x-ray crystallographic studies have recently been completed. The first must be viewed with some caution because the material used was reacted in frozen glasses (Land et al., 1982). The second study (Peckler et al., 1982) used an enantiomeric mixture of psoralen-thymine monoadducts. A unit cell containing three pairs of enantiomers was solved. The detailed information obtained is invaluable in modelling what the adduct structure is like in the helix and how it perturbs nearby bases. The angle between the psoralen and the thymine is clearly variable. Although helical constraints probably decrease the angle even more than crystal packing forces, a kink in the phosphodiester backbone will likely be present.

Using this information, extensive studies of DNA and RNA secondary structure have been done. While the secondary structure of cellular DNA is obvious without the aid of psoralen, the drug can be used to study the higher order structure (Wiesehahn et al., 1977; Hanson et al., 1976). The secondary structure of single-stranded viral DNAs is not quite so obvious. Hairpins have been mapped in fd virus relative to restriction enzyme cuts (Shen et al., 1979). The structures found appear to have functional significance.

DNA can also be crosslinked to RNA (Shen et al., 1977). This has been used to advantage in stabilizing R-loops for electron microscopy (Wittig and Wittig, 1979).

The reactivity of psoralen is much lower in RNA than in DNA (Isaacs et al., 1977), but some of psoralen's most important applications have come in examining RNA secondary structure. Crosslinks have been mapped in E. coli 16 S RNA (Wollenzein et al., 1979; Wollenzein and Cantor, 1982; chapter 6) and 5S RNA (Rabin and Crothers, 1979; chapter 5). Intermolecular associations between mRNA and 18S RNA (Nakashima et al., 1980), between U1 RNA and hnRNA in vivo (Calvet and Pederson, 1981), and between 18S RNA and 5S RNA (chapter 7) have been crosslinked by psoralen. The techniques needed to map crosslinks at high resolution are still being refined, but substantial progress is being made. Mapping of monoadducts is also possible (Youvan and Hearst, 1982; Bachellerie and Hearst, 1982). This does not yield as much information as crosslinks but is very useful, nonetheless.

The goal of the work presented in this thesis is to achieve a better understanding of RNA secondary structure. For the reasons stated above, psoralen crosslinking was chosen as the best means of attaining this. Clearly, knowledge about the nature of the reaction between psoralen and RNA has advanced greatly in recent years. In order to achieve a more rigorous understanding of how psoralen is affecting RNA secondary structure and to best take advantage of psora-

len's crosslinking abilities, a number of studies were done with model compounds to determine which features of RNA were important for reaction. Studies on the ability of different psoralen derivatives to crosslink RNA were also done. This work is followed by an analysis of the secondary structure of *D. melanogaster* 5S RNA and *E. coli* 16S RNA.

REFERENCES

Bachellerie, J. -P., Thompson, J. F., Wegnez, M. R., and Hearst, J. E. (1981) Nucl. Acids Res. 9, 2207-2222.

Bachellerie, J. -P. and Hearst, J. E. (1982) Biochemistry 21, 1357-1363.

Calvet, J. P. and Pederson, T. (1981) Cell 26, 363-370.

Cole, R. S. (1971) Biochim. Biophys. Acta 254, 30-39.

Dall' Acqua, F., Marciani, S., Rodighiero, G. (1970) FEBS Lett. 9, 121-123.

Dall' Acqua, F., Terbojevich, M., Marciani, S., Vedaldi, D., and Recher, M. (1978) Chem.-Biol. Interact. 21, 103-115.

Fox, G. E., and Woese, C. R. (1975) Nature 256, 505-507.

Glotz, C. and Brimacombe, R. (1980) Nucl. Acids Res. 8, 2377-2395.

Glotz, C., Zweib, C. Brimacombe, R., Edwards, K. and Kossel, H. (1981) Nucl. Acids Res. 9, 3287-3306.

Hancock, J. and Wagner, R. (1982) Nucl. Acids Res. 10, 1257-1269.

Hanson, C. V., Shen, C. -K. J., and Hearst, J. E. (1976) Science 193, 62-64.

Isaacs, S. T., Shen, C. -K. J., Hearst, J. E., and Rapoport, H. (1977) Biochemistry 16, 1058-1064.

Isaacs, S. T., Chun, C., Hyde, J. E., Rapoport, H., and Hearst, J. E. (1982) in Trends in Photobiology (Hélène et al., eds.), pp. 279-294, Plenum Publishing Corp.

Johnston, B. H., Johnson, M. A., Moore, C. B., and Hearst,

J. E. (1977) Science 197, 906-908.

Johnston, B. H., Kung, A. H., Moore, C. B., and Hearst, J. E. (1981) Biochemistry 20, 735-738.

Krauch, C. H., Kramer, D. M., and Wacker, A. (1967) Photochem. Photobiol. 6, 341-354.

Kanne, D., Straub, K., Rapoport, H., and Hearst, J. E. (1982a) Biochemistry 21, 861-871.

Kanne, D., Straub, K., Hearst, J. E., and Rapoport, H. (1982b) J. Am. Chem. Soc., in press.

Land, E. J., Rushton, F. A. P., Beddoes, R. L., Bruce, J. M., Cernik, R. J., Dawson, S. C., and Mills, O. S. (1982) J. Chem. Soc. Chem. Commun., 22-23.

Nakashima, K., Darzynkiewicz, E., and Shatkin, A. J. (1980) Nature 286, 226-232.

Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov. A. M., Stahl, D. A., Gupta, R. and Woese, C. R. (1981) Nucl. Acids Res. 9, 6162-6189.

Noller, H. F. and Woese, C. R. (1981) Science 212, 403-411.

Pathak, M. A., Kramer, D. M., and Fitzpatrick, T. B. (1974) in Sunlight and Man (ed. by Pathak et al.), pp. 335-368, Univ. of Tokyo Press, Tokyo.

Peckler, S., Kanne, D., Rapoport, H., Hearst, J. E., and Kim, S. -H. (1982) J. Mol. Biol., submitted.

Rabin, D. and Crothers, D. M. (1979) Nucl. Acids Res. 7, 689-703.

Rinke, J., Yuki, A., and Brimacombe, R. (1976) Eur. J. Biochem. 64, 77-89.

Ross, A. and Brimacombe, R. (1979) Nature 281, 271-276.

Scott, B. R., Pathak, M. A., and Mohn, G. R. (1976) Mutation Res. 39, 29-74.

Shen, C. -K. J., Hsieh, T. -S., Wang, J. C. and Hearst, J. E. (1977) J. Mol. Biol. 116, 661-679.

Shen, C. -K. J., Ikoku, A., and Hearst, J. E. (1979) J. Mol. Biol.127, 163-175.

Song, P. -S., Harter, M. L., Moore, T. A., and Herndon, W. C. (1971) Photochem. Photobiol. 14, 521-530.

Straub, K., Kanne, D., Hearst, J. E., and Rapoport, H. (1981) J. Am. Chem. Soc. 103, 2347-2355.

Tjerneld, F., Norden, B., and Ljunggren, B. (1979) Photochem. Photobiol. 29, 1115-1118.

Wagner, R. and Garrett, R. A. (1978) Nucl. Acids Res. 5, 4065-4075.

Wiesehahn, G. P., Hyde, J. E., and Hearst, J. E. (1977) Biochemistry 16, 925-932.

Wiesehahn, G. P. and Hearst, J. E. (1978) Proc. Nat. Acad. Sci. USA 75, 2703-2707.

Wittig, B. and Wittig, S. (1979) Biochem. Biophys. Res. Comm. 91, 554-559.

Wollenzein, P. L., Hearst, J. E., Thammana, P., and Cantor, C. R. (1979) J. Mol. Biol. 135, 255-269.

Wollenzein, P. L. and Cantor, C. R. (1982) J. Mol. Biol., in press.

Youvan, D. C. and Hearst, J. E. (1982) Anal. Biochem. 119, 86-89.

Zweib, C. and Brimacombe, R. (1980) Nucl. Acids Res. 8, 2397-2411.

Chapter Two:
Separation and Analysis of the Adducts Produced by Photoreaction of Psoralen with RNA

INTRODUCTION

In order to understand the details of reaction between psoralen and RNA, it is first necessary to have a reliable method of separating the products from the reactants. To this end, three different methods of separation, useful for different needs, have been developed. The ability to separate the different products has allowed us to study the kinetics of reaction as well as the types of products formed.

Paper electrophoresis at pH 3.5 was chosen to separate psoralen adducts because simultaneous separation of all four unmodified nucleotides can be achieved. This is useful when classical RNA sequencing is done (see chapter 5). This technique provides good separation of monoadducts from crosslinks with some separation of the different types of monoadducts. Paper electrophoresis at pH 8.8 provides much better separation of the different types of monoadducts but separation of unmodified nucleotides is lost. HPLC provides for greater resolution of the different stereoisomers of adducts and is thus preferable for use in detailed studies of structure.

Parts of this work were performed with Jean-Pierre Bachellerie and Maurice Wegnez, and appeared in Nucleic

Acids Research, volume 9, pp. 2207-2222, May 1981. HPLC and NMR work was performed with Stephen Isaacs.

MATERIALS AND METHODS

Materials. Homopolymers were obtained from P.L. Biochemicals. Nucleotide monophosphates and T2 RNase (E.C. 3.1.4.23) were purchased from Sigma. Calf intestine alkaline phosphatase was (E.C. 3.1.3.1) a Boehringer-Mannheim product. ^{3}H-HMT (3.7×10^{7} cpm/μg) was synthesized by Steve Isaacs of this lab according to Isaacs et al. (1982). Uniformly labeled ^{32}P-tRNA from cultured Drosophila cells was prepared as indicated in chapter 5.

Irradiation of Samples. The samples to be irradiated were put in Eppendorf tubes which were placed in a water bath. The bath was surrounded by a double-walled glass vessel containing a circulating 40% (w/w) cobaltous nitrate solution. This solution served as a temperature regulator and an ultraviolet and visible light filter with a maximum transmittance of 365 nm light (window: 320-380 nm). The samples were irradiated by two GE 400 W mercury lamps, one on each side of the sample. The light intensity at the surface of the sample was approximately 100 mW/cm^2. After irradiation, polymers were extracted twice with phenol and then ethanol precipitated two or more times to remove all traces of unreacted HMT. The RNA was redissolved in 50 mM sodium acetate, pH 5, containing 2000 units/ml T2 RNase. For up to 50 μg of RNA, 10 μl was used. Digestions were

done at 37°C for 18 hours. After photoreaction, mononucleotides were extracted two or more times with chloroform/isoamyl alcohol (24:1, v/v) and then electrophoresed directly.

Electrophoresis. Electrophoresis was done on Whatman #1 chromatography paper in the form of 46 x 57 cm sheets (VWR Scientific) or 27 cm wide rolls (Savant Instruments). The running buffer was 5% acetic acid, 0.5% pyridine (v/v/v), pH 3.5. Runs were done in a Savant LT 48 tank, normally for 50 minutes at 3000 V. Longer runs at 4000 V for 3 or more hours were also done. An aliquot of a mixture of (^{32}P)3'-phosphate mononucleotides was added to each sample to serve as mobility markers. Electrophoresis at pH 8.8 was done with a buffer containing 80 g Tris base and 8 ml acetic acid per liter. After electrophoresis, the paper was dried and cut into 0.5 or 1 cm slices. In order to obtain high counting efficiencies, 0.5 ml of water was added to each slice for one hour to elute the nucleotides from the paper. Each fraction was counted after addition of a 5 ml scintillation mixture containing 2 parts toluene, 1 part Triton X100, 4 g/l omnifluor (98% PPO, 2% Bis MSB).

When it was necessary to recover the electrophoresed samples, the paper strips were counted directly at low efficiency in toluene with 6 g/l PPO and .075 g/l POPOP. After counting, the paper strips were rinsed with toluene and dried. They were then each placed in a 200 μl pipet tip and rinsed with 200 μl of H_2O in order to elute the

nucleotides from the paper.

Photoreversal. Photoreversal of crosslinks and monoadducts was done with a hand-held Rayonet 6 W short wave UV lamp. The lamp was positioned 5 cm above the sample to be reversed. The optical density of the samples was low enough that the rate of photoreversal was independent of concentration.

HPLC. Samples run on HPLC were digested in two ways. Preparative scale samples digested with 1 M HCl at 60° for 4 hours were then stripped down and resuspended in Tris (adjusted to pH 7.2). BAP (bacterial alkaline phosphatase, Sigma, grade VII) was added, 100 units per addition, and then incubated at 37°. Aliquots were monitored by paper electrophoresis at pH 3.5 to see when dephosphorylation was complete. Analytical scale samples (less than 1 mg RNA) were enzymatically digested. Samples were diluted to 1 ml, 100 mg of crude snake venom phosphodiesterase (Sigma) and 100 units of BAP added, followed by incubation at 37°. The degree of reaction was followed by electrophoresis.

After digestion to nucleosides, samples were applied to a C-18 Sep Pak cartridge (Waters). This was eluted with H_2O, 50 mM $NaPO_4$ (pH 2.1), H_2O, and CH_3OH. The CH_3OH fractions were combined and concentrated. Samples were injected onto a 10 mm x 25 cm reverse phase C-18 Ultrasphere ODS column (Altex-Beckman). Flow rates were 4 ml/min with H_2O/CH_3OH gradients. Optical density of all

samples was monitored at either 250 or 330 nm and fractions were taken and counted for ^{3}H. ^{1}H NMR spectra were done at 360 MHz on a Nicolet Technologies NT360 NMR spectrometer. Sample preparation and NOE experiments were done as by Straub et al. (1981).

RESULTS

The photoproducts of HMT with different types of natural and synthetic RNAs have been studied, taking advantage of a highly radiolabeled psoralen derivative synthesized in this laboratory (Isaacs et al., 1982b). After the photoreacted RNA has been completely hydrolyzed by extensive digestion with T_2 RNase, the products can be separated by electrophoresis on neutral paper at pH 3.5.

Reaction of [^{3}H]HMT with all the natural RNAs we have tested so far yields only two major labeled peaks separable by paper electrophoresis. Typical radioactivity profiles of RNA hydrolysates are shown in Figure 1, for a mixture of E. coli tRNAs and for poly(U). The relative mobilities of each peak (taking $m^{7}G$ as a reference, $R_F(m^{7}G)=-1$) are .62 and .99 for peak 1 and peak 2, respectively. In every case, we observed that the (^{3}H) cpm ratio between peaks 1 and 2 was highly dependent upon the light dose. The proportion of peak 2 photoadducts increases with irradiation time (as shown here in Figure 1b for poly(U)). Moreover, "pulse-chase" experiments, in which a short irradiation of RNA in the presence of [^{3}H]HMT is followed by removal of unreacted psoralen and further irradiation, have indicated that a

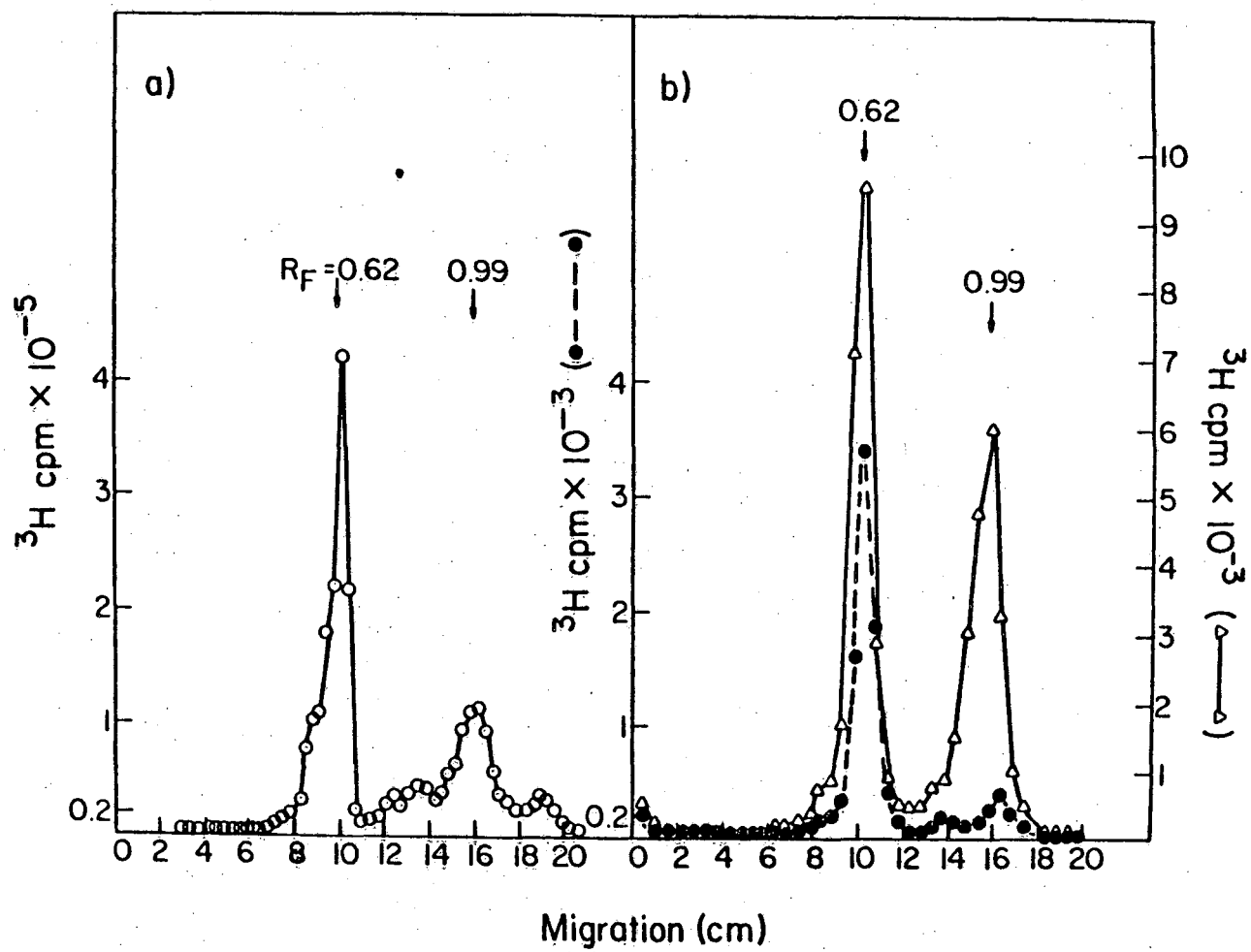

a)
b)
R_F = 0.62
0.99
0.62
0.99
^{3}H cpm × 10^{-5}
^{3}H cpm × 10^{-3} (●——●)
^{3}H cpm × 10^{-3} (△——△)
Migration (cm)

Figure 1: Paper electrophoretic analysis of T_2 RNase digests of HMT-photoreacted E. coli tRNAs and poly(U). RNA was photoreacted with [^{3}H]HMT, extracted with phenol, and ethanol precipitated. The samples were then hydrolyzed with T_2 RNase and analyzed by paper electrophoresis at pH 3.5. A mixture of [^{32}P] labeled ribonucleotide monophosphates were comigrated as markers. Radioactivity profiles were determined by liquid scintillation counting of 0.5 cm paper slices. (a) A mixture of E coli tRNAs (200 μg/ml) was photoreacted with 1 μg/ml [^{3}H]HMT for 10 min at 5° in 5 mM NaCl, 1 mM Tris-HCl (pH 7.5), and 1 mM EDTA. (b) Poly(U) (660 μg/ml) was photoreacted with 1.3 μg/ml [^{3}H]HMT at 16° in 1 mM Tris-HCl (pH 7.6) and 1 mM EDTA for 15 s (•- -•) or 2 min (△—△).

fraction of the pulse-labeled peak 1 component can be converted into peak 2 product by 320-380 nm light during the chase period (not shown). In order to aid in the identification of the composition of these two major peaks, the reaction was examined in simple systems such as homopolymers and mononucleotides.

When HMT was photoreacted with the four ribohomopolymers in standard conditions, the reactivity of poly(U) was found to be 10-20 times higher than that of other synthetic homopolymers (Table I). The paper electrophoresis analysis of T_2 RNase digests showed that only poly(U) yields a profile similar to that observed with natural RNAs. The much less efficient reaction of HMT with poly(C) gives rise to a labeled component with the same mobility (R_F= 0.14) as a minor component seen in natural RNA, as well as a component (R_F = 0.62) identical to the one observed with poly(U). These are detected in varying proportions, depending upon how the photoreacted polymer is handled (see Discussion). A more complex electrophoretic pattern of labeled products was obtained for poly(G) (not shown). Most of the products remained at the origin when electrophoresed at either pH 3.5 or 8.8. At pH 3.5, minor products formed which had both positive and negative mobilities.

The reaction of HMT with the different mononucleotides was also studied. The products detected with the mononucleotides have mobilities identical to those observed in the respective homopolymers. The reactions were found not to

Table I: Photoreactivity of ribohomopolymers with HMT*

Polynucleotide	Relative Photoreactivity
Poly(U)	1
Poly(C)	.064
Poly(G)	.112
Poly(A)	.058

*Homopolymers were reacted as in Figure 1 except irradiations were for 2 min.

vary with temperature in the range 4-50°C. Ionic concentration did have an effect on the reactivity. Raising the concentration of NaCl from 0 mM to 100 mM increased the amount of HMT incorporated into UMP by 25%. In the case of mononucleotides, the photoreactivity of uridine was higher than that of the other major bases. CMP was about one third as reactive while no reactivity above backround could be detected for AMP and GMP. We also examined the HMT photoreactions with different uridine analogs, in the mononucleotide monophosphate form. No reaction could be detected with dihydrouridine, even though it would have been possible to observe a reaction which is 20 times less efficient than the reaction with uridine. Pseudouridine and 4-thiouridine were approximately 10 times less reactive than uridine.

The electrophoretic mobilities of the observed HMT-mononucleotide products are listed in Table II. These mobilities have been compared to those calculated from an empirical equation (Sommer, 1979). Based on this and on the nature of the reactive nucleotides used in the simple systems, a tentative assignment can be made as to the identity of these products. The major peak (peak 1, R_F = 0.62) in Figure 1 has the same mobility as predicted for an HMT monoadduct of uridine monophosphate, while the other major product (peak 2, R_F = 0.99) has the mobility expected for an HMT crosslink between two uridine monophosphates. For this reason, as well as additional evidence shown below, we shall refer to these products as U monoadduct and U-U crosslink,

Table II: Electrophoretic mobilities of photoproducts at pH 3.5.

Compound	Predicted R_F[+]	Observed R_F
CMP	.21	.19
AMP	.33	.37
GMP	.76	.74
UMP	.93	.93
UMP + HMT	.64	.57
		.61
		.66
UMP + UMP + HMT	.95	.99
		1.04
UMP + CMP + HMT	.58	.57
CMP + HMT	.14	.14

[+]The predicted mobilities were calculated from the equation $R_F = 45.4Q/M^{2/3}$ where Q is the net charge and M is the molecular weight (Sommer, 1979).

respectively. When the two components are dephosphorylated by alkaline phosphatase, they both remain at the origin during electrophoresis at pH 3.5. This is exactly what is predicted by their assigned structures.

The stability and reactivity of the presumptive U monoadduct and U-U crosslink were also examined. Both products were purified by paper electrophoresis and assayed in various conditions. A further irradiation at 360 nm resulted in no detectable change in the electrophoretic profiles. When irradiation of the U-U crosslink at 360 nm was carried out in the presence of additional nucleotides, the U-U crosslink was again unchanged. The U monoadduct, on the other hand, could be converted to a U-U crosslink if the irradiation was carried out in the presence of additional UMP. In the presence of additional CMP, the irradiation of U monoadduct gave rise to a new product migrating as would be predicted for a U-C crosslink using the empirical equation mentioned above (Figure 2). However, even with careful handling, this new product slowly converts to a compound migrating like the U-U crosslink (see Discussion). No new product was detected when the further irradiation of U monoadduct was performed with additional AMP or GMP. Both of these crosslinks could also be photoreversed (Figure 3). After a 260 nm irradiation of (^{3}H) HMT-uridine monoadduct, the fraction of tritium counts remaining at the origin when electrophoresed on paper increased proportionally with the light dose (not shown). When irradiated with short UV

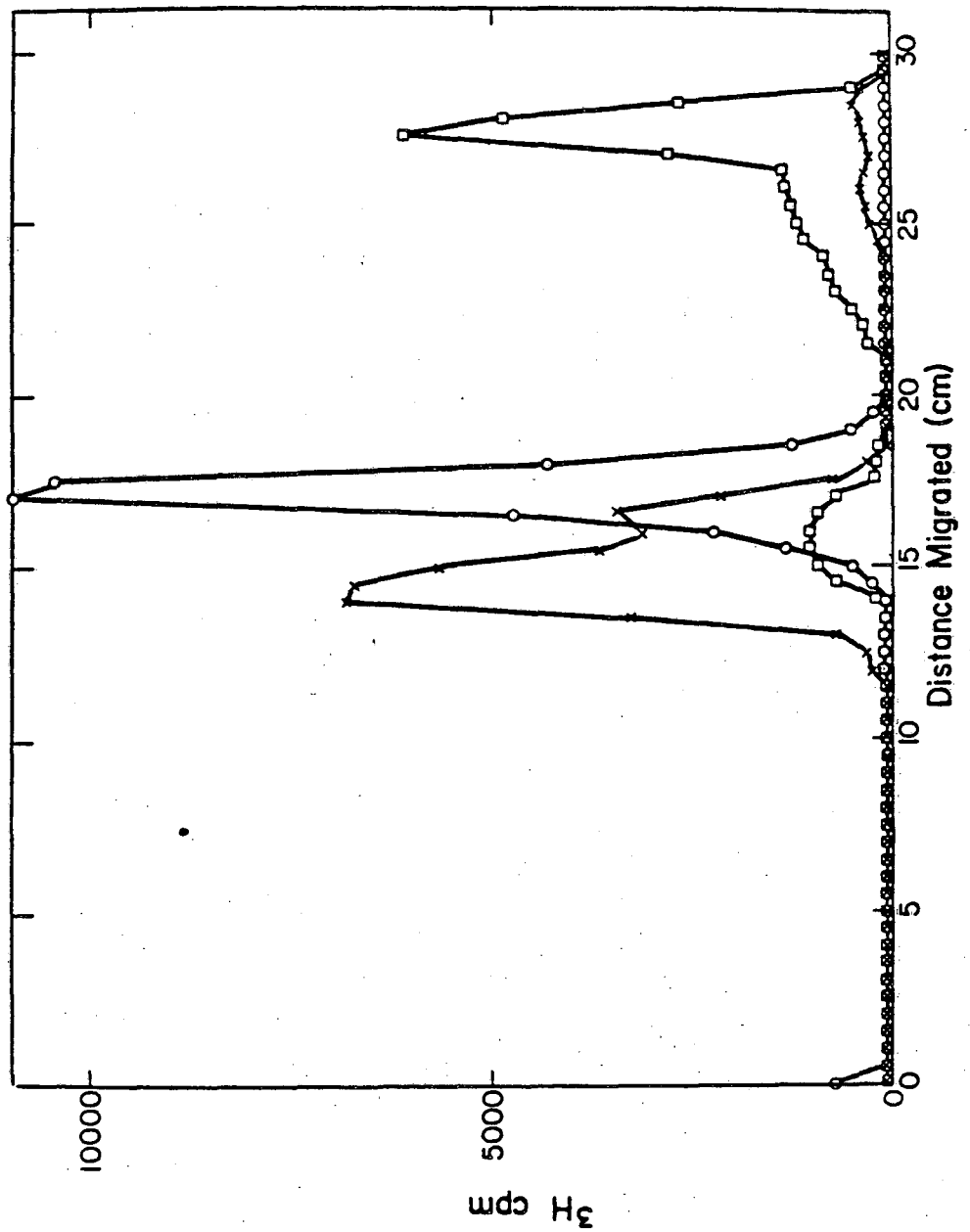

^3H cpm
10000
5000
0
0
5
10
15
20
25
30
Distance Migrated (cm)

Figure 2: Formation of crosslink from monoadduct. U monoadduct from E. coli tRNAs was purified and reacted further. A 10 min irradiation was performed in 1 mM Tris-HCl (pH 7.4) 0.1 mM EDTA at 20°. The concentration of monoadduct was 2.4×10^{-7} M and no additional nucleotides (o—o), 50 mg/ml UMP (□—□), or 50 mg/ml CMP (×—×) was added.

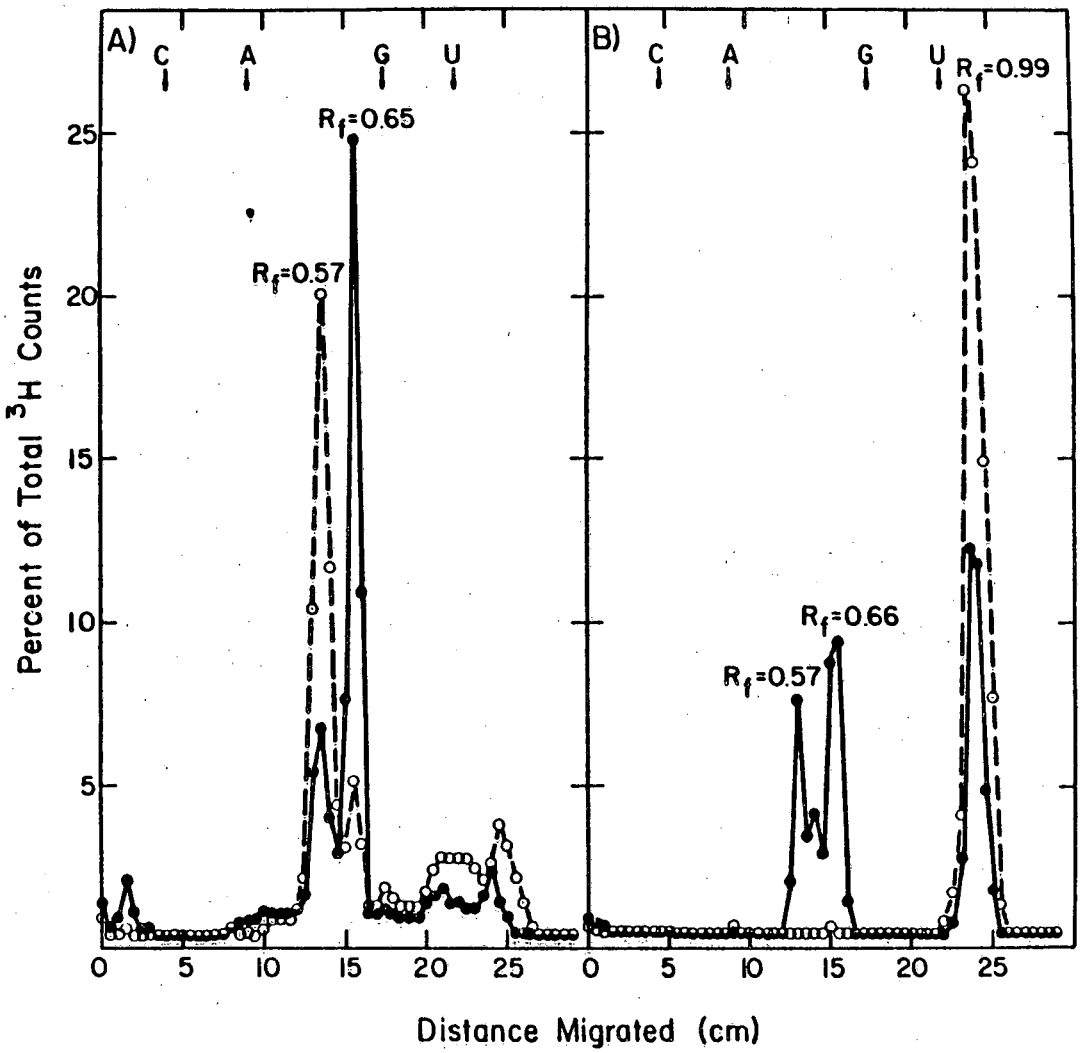

A)
C
A
G
U
R_f=0.65
R_f=0.57
B)
C
A
G
U
R_f=0.99
R_f=0.66
R_f=0.57
Percent of Total ^3H Counts
0
5
10
15
20
25
Distance Migrated (cm)

Figure 3: Photoreversal of crosslinks. The U-C crosslinks (A) and the U-U crosslinks (B) formed as shown in Figure 2 were purified and then reversed with 20 minutes of short-wave ultraviolet light. The crosslinks are shown before (o--o) and after (•—•) reversal.

light, the uridine-HMT-uridine crosslink was progressively converted into a labeled compound migrating like the uridine monoadduct (Figure 3b). This was, in turn, broken down to a labeled product which remains at the origin after further UV exposure (not shown). The 260 nm irradiation of the presumptive U-C crosslink resulted in the transient detection of U monoadduct and a small amount of C monoadduct (Figure 3a).

The kinetics of formation of these products was studied. In order to get efficient reaction, large excesses of nucleotides have to be added. A plot of these data is shown in Figure 4. The straight lines obtained indicate a reaction which is first order in monoadduct. In order to determine the order of reaction in which monoadduct is formed, the change in the initial rate of reaction was observed with varying concentrations of UMP. This is shown in Figure 5.

In order to confirm the presumptive nature of the HMT-uridine photoproducts, further experiments were done using ^{32}P labeled RNA. In vivo uniformly labeled ^{32}P-tRNA was photoreacted with ^{3}H-HMT and then hydrolyzed by T_2 RNase before paper electrophoresis. All of the tritium labeled peaks are listed in Table 3. The U monoadduct could actually be resolved into three components, while the U-U crosslink was separated into two species. The tritium to ^{32}P ratios also support the assignments. The crosslink has twice as much ^{32}P per HMT when compared to the monoadduct. All of these species were also reversed with short UV light.

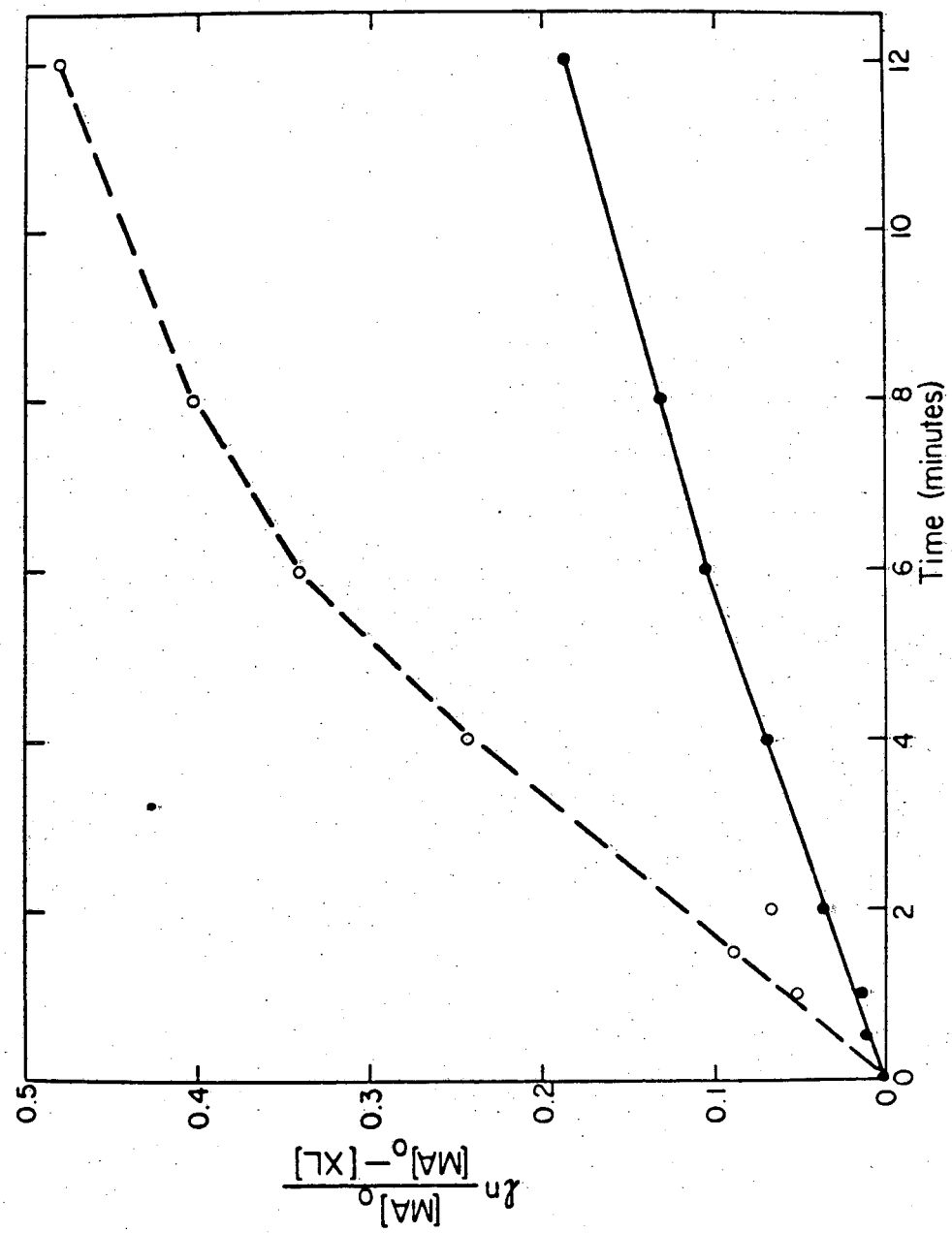

$\ln \frac{[MA]_o}{[MA]_o - [XL]}$
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
Time (minutes)

Figure 4: Kinetics of formation of crosslink from monoadduct. The formation of the U-U crosslink (○···○) and the U-C crosslink (●—●) as described in Figure 3 is shown in a first order plot. The initial concentration of monoadduct is $[MA]_0$ and the concentration of crosslink at any time is $[XL]$. The initial concentration of monoadduct was 2.4×10^{-7} M and the initial concentration of mononucleotides was 50 mg/ml.

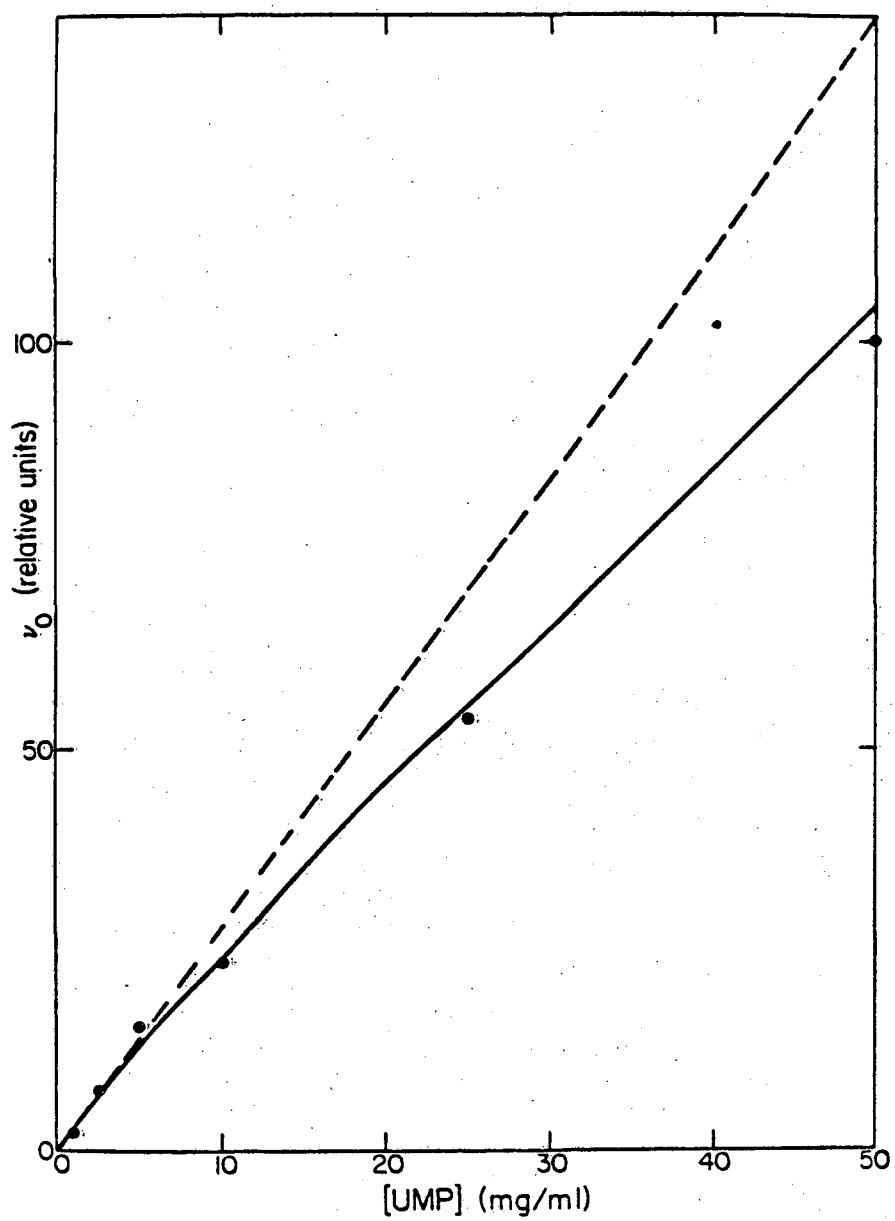

100
50
0
v_o (relative units)
0
10
20
30
40
50
[UMP] (mg/ml)

Figure 5: Kinetics of formation of monoadduct. Varying concentrations of UMP were reacted with 15 µg/ml HMT, 1 mM Tris-HCl (pH 7.4), and 0.1 mM EDTA at 20° for 30 sec and analyzed by paper electrophoresis. The curved lines were generated from a best fit of the data to an exponent in the equation listed in the DISCUSSION. The dashed line is the best fit for the four lowest concentrations while the solid line is for all concentrations.

Table III: Photoproducts of reaction between [^{3}H]HMT and [^{32}P]tRNA.

R_F	% of Total ^{3}H	^{3}H/^{32}P	Products of UV Photoreversal
.57	4.48	28	UMP, HMT
.61	24.22	33	UMP, HMT
.66	56.88	35	cUMP, HMT
.98	6.61	18	UMP, cUMP, U monoaiducts (R_F = .57, .66), HMT
1.04	2.61	-	cUMP, U monoadduct (R_F = .66), HMT

The remaining 5.2% of the tritium was present in HMT breakdown products and incompletely hydrolyzed material. The RNA was irradiated for 10 min in a solution of 1 µg/ml HMT, 5 mM NaCl, 1 mM Tris-HCl (pH 7.4), and 0.1 mM EDTA.

When the U monoadducts were reversed (Table III), the slower two yielded UMP while the faster one yielded a product migrating like cyclic UMP. When additional UMP was added to the two faster migrating monoadducts (and then irradiated at 360 nm), U-U crosslink was formed. This was not performed on the slowest moving component (R_F = 0.57). When the crosslinks were reversed, monoadducts, UMP, and HMT were formed.

More resolution can be obtained in the separation of the U monoadduct by carrying out the electrophoresis for a longer time. A typical profile is shown in Figure 6a. The broad monoadduct peak is actually resolved into three subspecies with R_F = .57, .61, and .66. The relative proprotions of the components is highly dependent upon the ratio of the concentration of monoadduct to the concentration of T_2 RNase present during the hydrolysis of the photoreacted RNA. When the ratio is high (low enzyme concentration), the peak migrating with R_F = .66 is the major species. When the T_2 RNase concentration is raised sufficiently, this peak disappears with a concomitant increase in the size of the peak with R_F = .61. The faster migrating peak can also be quantitatively converted into the component migrating with R_F = .61 by a one hour treatment at pH = 1 (Figure 6b). The R_F = .66 component is resistant to dephosphorylation with calf alkaline phosphatase, unlike the R_F = .61 component (Figure 6c). The acid treatment of the R_F = .66 component allows its subsequent dephosphorylation by alkaline

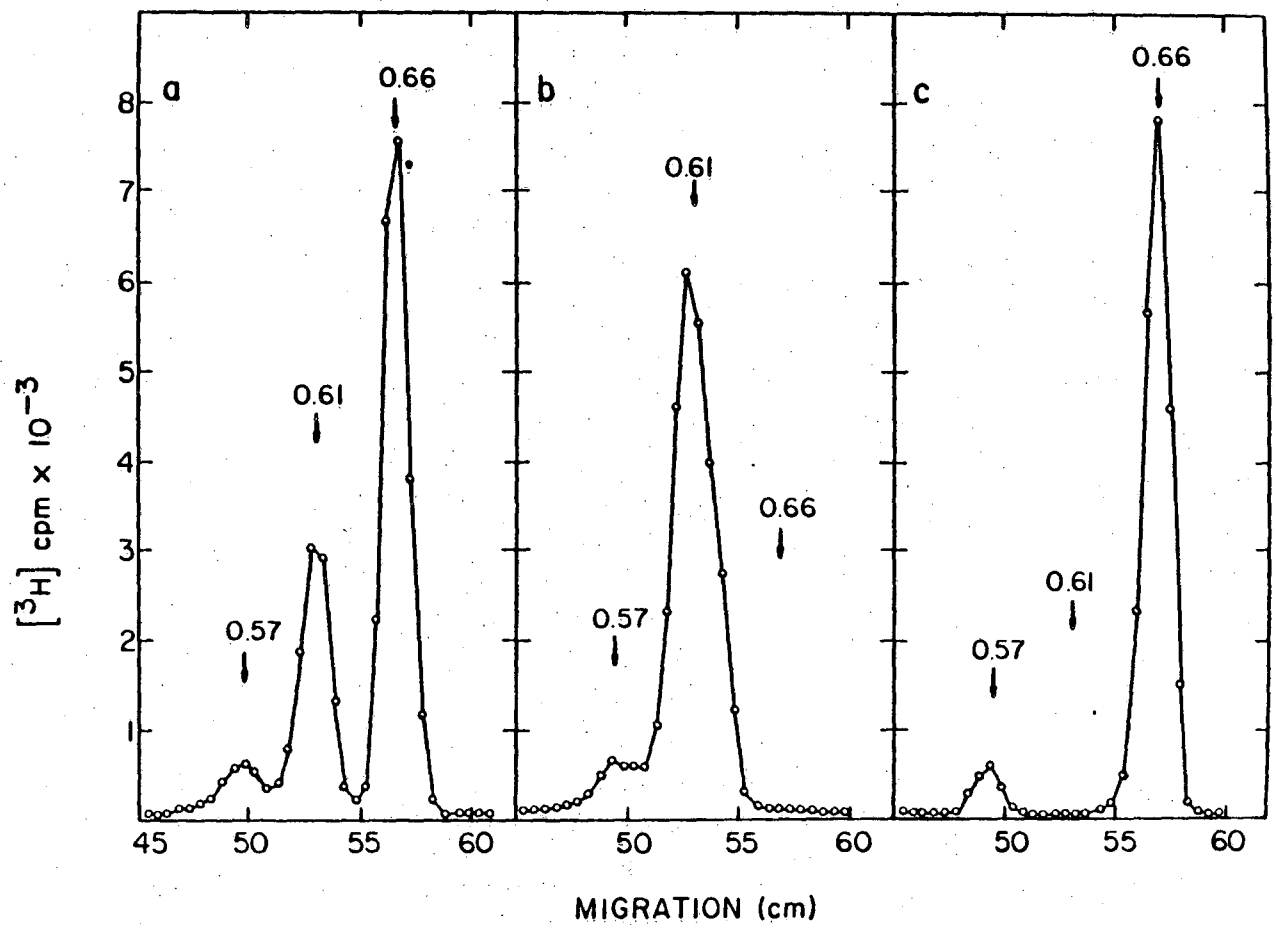

a
b
c
0.66
0.61
0.57
0.66
0.61
0.57
0.66
0.61
0.57
[3H] cpm x 10−3
MIGRATION (cm)
45
50
55
60
1
2
3
4
5
6
7
8

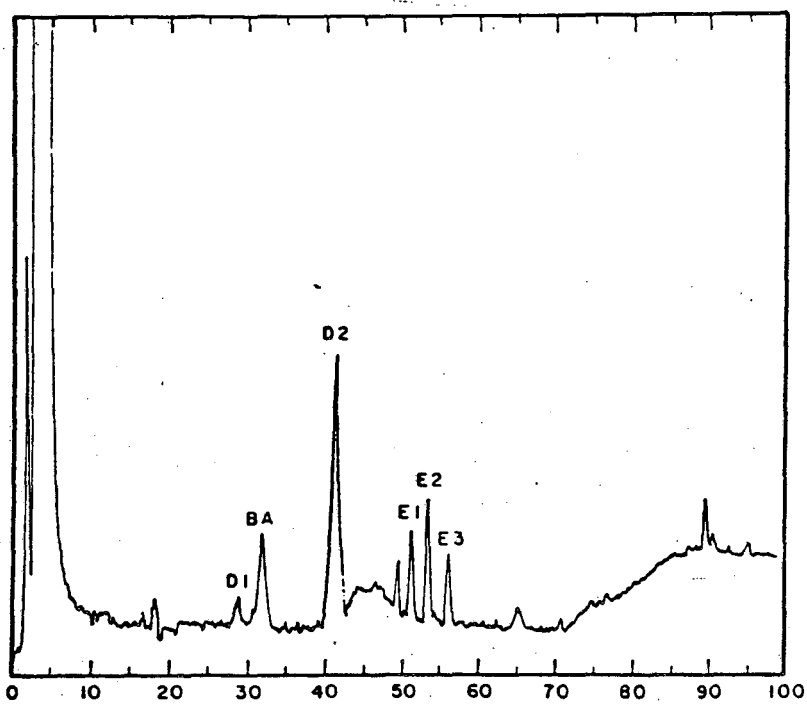
Figure 6: High resolution electrophoretic separation of UMP-HMT monoadducts. RNA containing HMT monoadducts was digested with T_2 RNase and electrophoresed at pH 3.5 on 90 cm sheets of paper. (a) No further treatment (b) The T_2 RNase digest was treated with 0.1 M HCl for 1 hr before electrophoresis. (c) The T_2 RNase digest was treated with calf intestine alkaline phosphatase (10 mUnits, 75 min, 37° in 50 mM Tris-HCl, pH 8.3) before electrophoresis. Each sample contained 6 picomoles of monoadduct.

phosphatase (not shown). It should also be noted here that the experiment reported in Figure 1 was carried out with a low substrate to enzyme ratio. This results in almost complete absence of the R_F = .66 component.

The properties of the R_F = .57 component are more complex. It can interconvert with a portion of the R_F = .61 component and is susceptible to phosphatase. When eluted from the paper, there is no absorption above 280 nm (lower wavelengths are obscured by contaminating fluors) while the R_F = .66 component has a spectrum characteristic of a furan monoadduct. Greater than 70% of the R_F = .66 component can be driven to crosslink after addition of nucleotide and reirradiation while less than 5% of R_F = .57 can.

These species were further examined by electrophoresis at pH 8.8. All mobilities at pH 8.8 will be denoted as R_F* to avoid confusion and are relative to the mobility of xylene cyanol. The R_F = .66 component migrates with R_F* = .53 which, after acid treatment, migrates at R_F* = .82. The R_F = .57 component yields an R_F* = 1.06 species. Also observed at pH 8.8 is an R_F* = .91 product which can be converted to R_F* = 1.06 by acid treatment. Crosslinks migrate at R_F* = 1.27.

HPLC profiles of the hydrolysis products of poly(U)·poly(A) irradiated at 5 mM NaCl is shown in Figure 7. Acid hydrolysis (Figure 7a) generates more products than enzymatic hydrolysis (Figure 7b). The major products


D2
BA
E2
E1
E3
D1
0
10
20
30
40
50
60
70
80
90
100

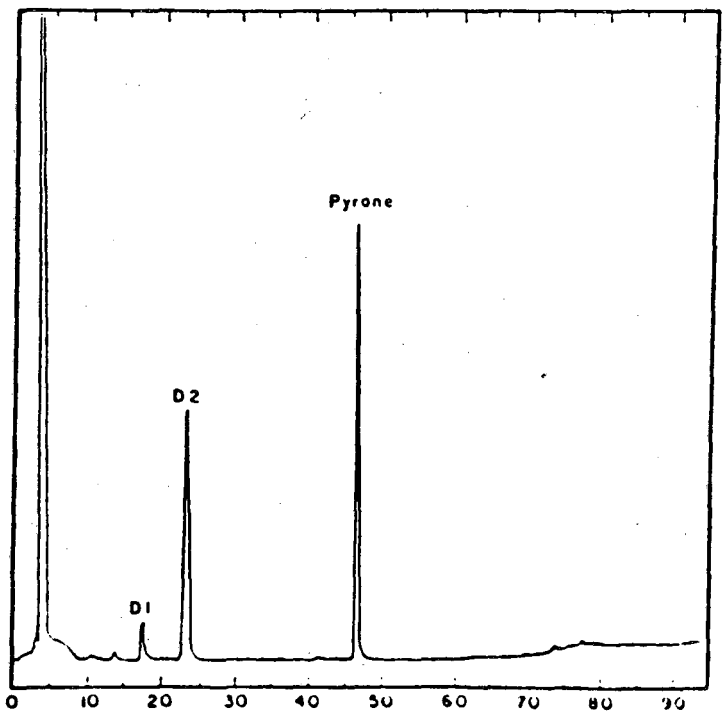

Pyrone
D2
D1
0
10
20
30
40
50
60
70
80
90

Figure 7: HPLC profile of HMT adducts. a) Acid digestion of poly(U)·poly(A) yields the expected furan diastereomers, base adduct, and a number of equilibrium products. The relative amounts of these are dependent on length of hydrolysis. b) Enzymatic digestion of a similarly prepared sample yields only the furan diastereomers and pyrone adduct. This sample was run in 10 mM sodium phosphate (pH 2.2) to keep the open coumarin adduct uncharged. The two samples were run on different columns so the retention times are not equivalent.

have been analyzed by NMR, CD, and UV absorption. D1 and D2 have identical CD, UV, and fluorescence spectra and these are identical to the diastereomeric furan monoadducts characterized by Straub et al (1981). NMR and NOE results confirm this equivalence. NMR spectra are shown in Figure 8. The base adduct (BA), in which the glycosidic bond has been cleaved, has been similarly characterized.

Both BA and the equilibrium products (E1, E2, etc.) are formed in increasing amounts when acid hydrolysis is used. Enough material has been isolated from some of the equilibrium products for NMR and UV spectra. It is clear that they arise from the furan adducts and that no alteration has been made to the coumarin moiety. There are differences in the NMR spectra (compared to normal furan adducts) but contaminants in the samples make it difficult to conclusively determine the spectra. The equilibrium products slowly revert to the furan adducts upon prolonged (greater than 1 month) sitting at room temperature in H_2O/CH_3OH. Conversely, the furan adducts can slowly convert to equilibrium products. This process is catalyzed by acid and is not dependent on CH_3OH. The coumarin adducts and crosslinks are present in low amounts and could not be analyzed directly. Their mobilities are what would be expected, based on the results of Straub et al. (1981).

DISCUSSION

Identification of photoproducts (monoadducts vs. crosslink). The process of identifying the photoproducts of the

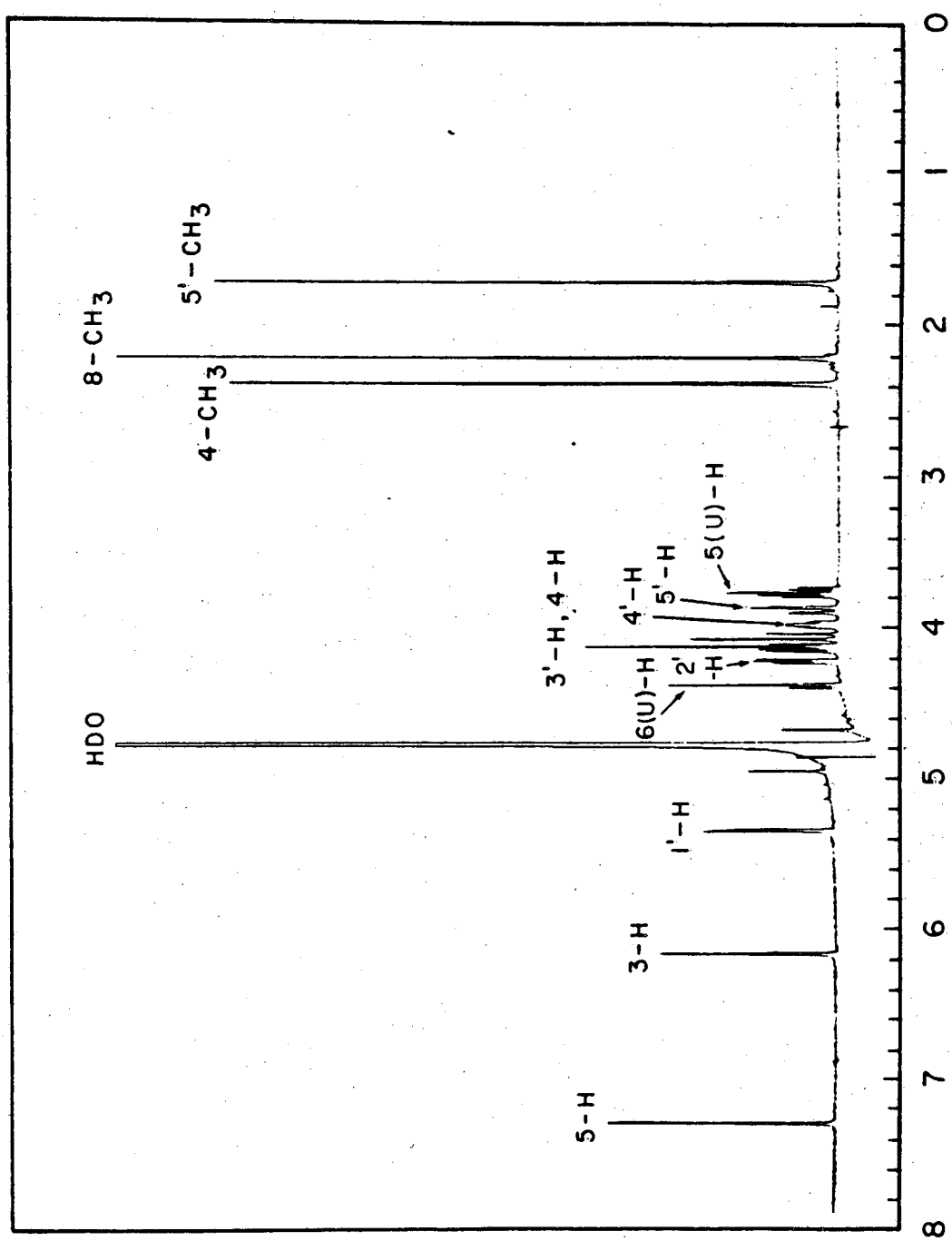

HDO
8-CH3
5'-CH3
4-CH3
3'-H, 4-H
6(U)-H
2'-H
4'-H
5'-H
5(U)-H
1'-H
3-H
5-H
0
1
2
3
4
5
6
7
8

Figure 8: NMR spectrum of diastereomer two from poly(U)-poly(A).

reaction between HMT and natural RNA is simplified greatly by the use of model compounds such as homopolymers and mononucleotides. In the case of both polymers and nucleotides, uracil is the most reactive of the four bases. For uridine, there is no difference in reactivity between the nucleotide and the polymer. This is in agreement with previous work (Pathak et al., 1974; Krauch et al., 1967) and indicates that polymerization has no effect on the reaction. The products obtained from poly(U) and UMP are identical in electrophoretic mobility to the major products obtained in natural RNAs. This suggests that the reaction in natural RNA involves principally uridine.

This assignment is supported by the use of an empirical equation which can be used to predict electrophoretic mobilities on neutral paper. This equation assumes that only the charge and molecular weight of a compound determine its mobility. While other factors are certainly involved, the equation provides a good first approximation of the mobility of the common and modified nucleotides that have molecular weights from 200 to 2000 D. We have used the mobility of UMP as our standard and equated its R_F with that observed by Sommer (1979), who measured mobilities relative to m^7G. Using the assumption that, at pH 3.5, addition of HMT does not affect the net charge of the nucleotide to which it is attached, the mobilities were calculated and compared to those observed. The close agreement is strong evidence for the assignments made. That the global charge of the uridine

adducts does remain identical to that of unreacted uridine (i. e. R_F = 0 at pH 3.5) was experimentally confirmed by the absence of mobility of the dephosphorylated photo-products as well as the base adduct (no sugar or phosphate).

The possibility that the peak 2 product results from an incomplete hydrolysis of the phosphodiester chain by T_2 RNase was examined. The nucleolytic activity of various RNases can actually be altered adjacent to the photoadduct. Resistant dinucleotide diphosphates containing a U-monoadduct can be detected for shorter periods of T_2 RNase digestion, and these have also been characterized by paper electrophoresis (Bachellerie and Hearst, 1982). An HMT monoadduct of UpUp migrates only marginally slower than the crosslink between two uridine monophosphates (R_F = .93 instead of .99). However, these compounds can be easily distinguished after dephosphorylation by BAP (R_F = .50 for UpU monoadduct, R_F = 0 for U-U crosslink), thus providing a good assay for monitoring the extent of T_2 RNase digestion. In the experiments described here, no detectable fraction of resistant phosphodiester linkages was observed. Another possibility, pUp-HMT, migrates significantly faster than peak 2.

Photoreversal (first observed by Musajo et al., 1967) of peak 2 yields additional evidence for this assignment. Peak 2 can be completely converted to monoadduct which is in turn reversed. When ^{32}P crosslink is reversed, the ^{32}P is ultimately found as UMP and cUMP. If peak 2 contained partially hydrolyzed monoadduct, reversal would lead immediate-

ly to HMT and other ^{32}P products.

Kinetics of peak 1 and peak 2 appearance were also studied during the initial period of 360 nm irradiation, when changes in HMT concentration due to photoreaction and breakdown can be neglected. While the rate of peak 1 appearance is linear with light dose, the rate of peak 2 production is a quadratic function of light dose. This is predicted by the previous observation that crosslink formation is a two photon event (Johnston et al., 1977).

The preferential reactivity with uridine agrees with the observations of Ou and Song (1978). They found that the major psoralen photoproduct in fluorouracil substituted tRNA involved fluorouracil. Also, the absence of reactivity observed for dihydrouridine agrees well with the hypothesis that the 5, 6 double bond of the pyrimidine is involved in the psoralen addition. This process is also clearly inhibited when this double bond is sterically hindered by a C_6-C_1 ribose linkage as in the case of pseudouridine.

Besides uridine photoadducts, some minor products have been detected. Their mobilities agree well with predicted values (Table II): a CMP-monoadduct was detected by irradiating CMP, poly(C), or natural RNA and a U-C crosslink was detected by irradiating U monoadduct in the presence of CMP. Identification of the C-photoproducts was complicated by the fact cytidine deaminates readily when the 5, 6 double bond is saturated by the addition of HMT. This process is accelerated by heat but it occurs even at low temperature and

neutral pH. CMP monoadduct is then converted to UMP monoadduct. Deamination has been observed previously with psoralen adducts (Musajo et al., 1967), cytidine dimers (Liu and Yang, 1978), and bisulfite adducts (Slae and Shapiro, 1978). This process clearly interferes with determining the accurate extent of cytidine photoaddition in natural RNA. In DNA, this problem can be solved because uridine adducts arising from cytidine deamination can be separated from thymidine adducts (Straub et al., 1981). The HMT adduct to guanosine was not obsereved in native RNA, either because of its low yield or its complexity.

The exact nature of the guanine adducts was not elucidated because of the number and low yield of products. The fact that, at pH 3.5, most of the products had no mobility, suggests that the HMT reaction introduces a positive charge on the guanine ring. This would be the case if the 7, 8 double bond reacted, quarternizing the N7. This compound would be unstable at a pH greater than 7 (Townsend and Robin, 1963) and thus explain the difficulty in reproducing exact product profiles. The glycosidic bond in such compounds is also very labile (Lawley and Brookes, 1963) and, if broken, would produce similar results. The principal minor product of the guanine reaction has a mobility (R_F = .52) virtually identical to that predicted for a simple addition of HMT to GMP with no alteration in charge. This could occur by addition to the 6 carbonyl bond, for example.

There are three other reports in the literature of

psoralen covalently reacting with a purine, two involving adenine and one involving inosine. Ou and Song (1978) used paper chromatography to separate the photoadducts formed in tRNA. Anything which did not leave the origin was assumed to be photoadduct. The composition of the photoadduct was determined by subtracting the amount of the nucleosides found in the photoreacted sample from that of a control sample. This type of indirect proof is not sufficient, especially since it has been determined that the presence of a psoralen adduct can markedly inhibit the rate of enzymatic digestion (Bachellerie and Hearst, 1982). It is more likely that adenine was simply adjacent to the photoadducts, was not cleaved by the enzymes, and remained at the origin. Krauch et al. (1967) noted a low level of reaction of psoralen to poly(A). Unfortunately, a critical control experiment was omitted so it is not certain whether the observed binding is real. These experiments employed Sephadex G100 to separate unreacted psoralen from the polymers. The control where psoralen was irradiated, added to the polymer, and purified was not done. Hochkeppel and Gordon (1979) reported addition to inosine; and, while their results did not directly show the presence of an inosine adduct, their evidence for a crosslink between poly(I) and poly(C) is plausible in light of the poly(G) reactivity.

Examination of the putative intercalation complex would seem to indicate that the 7,8 double bond would be most favorably aligned. The difference in reactivity among the

purines would then be explained by differences in the π electron structure or by the solution structure of the polymers which allow differential access to the photoreactive bonds. The lack of reactivity at the nucleotide level tends to support the latter conclusion. Outside binding of psoralen to the helix is unlikely as the sole explanation since poly(I)·poly(C) can be crosslinked.

None of these possibilities explains the unusual behavior of poly(G) in the presence of added salt. Unlike the addition of HMT to other polymers (chapter 3), the incorporation of HMT increases by a factor of two as the concentration of NaCl is raised from 0 to 100 mM. Also surprising is the fact that addition of poly(C) (in conditions where the polymers should base pair) causes no change in the amount of HMT added. While the answers to these questions are no doubt interesting, these problems have not been pursued because of their unimportance when considering native RNA.

Assignment of isomers of U-monoadduct and U-U crosslinks.

The number of possible isomers from the reaction of HMT with any of the bases is very large. Clearly, most of these would not be separable by paper electrophoresis. Nevertheless, high resolution migration allowed the detection of three subspecies in the peak I (U monoadduct) component obtained either from natural RNAs or U-containing synthetic polymers. However, only two of these subspecies can be detected after photoreacting with UMP. The reasons for an additional fast-migrating (R_F = 0.66) U monoadduct in

the case of polynucleotides can be understood upon considering the RNA hydrolysis process itself. T_2 RNase digestion involves the formation of a 2'-3' cyclic phosphate intermediate. The decyclization of this into the final 3' phosphate nucleotide is also catalyzed by the enzyme. Cyclic phosphate mononucleotides migrate slightly but significantly faster than the noncyclic form (not shown). Our results are consistent with the fastest moving U monoadduct (R_F = 0.66) being a 2'-3' cyclic phosphate nucleotide, the conversion of which is strongly inhibited by the presence of HMT. As expected, increased amounts of T_2 RNase promotes the conversion of R_F = 0.66 species into R_F = 0.61 species. Also, the classical procedure of using acid to decyclize the phosphate (Sanger et al., 1965) allows us to perform the same conversion. This assignment is in line with the inhibition of dephosphorylation of the fast moving U monoadduct (R_F = 0.66). The dephosphorylation becomes possible once the decyclizing acid-treatment has been carried out. The short-UV light reversal experiment reconfirms this identification, since the R_F = .61 U-monoadduct yields (^{32}P) labeled noncyclic UMP and the R_F = .66 monoadduct yields (^{32}P) labeled cyclic UMP (Table III).

The pH 8.8 electrophoresis data shed additional light on the assignment of the monoadduct subspecies. At pH 8.8, the 3' phosphate will contain two negative charges while the 2', 3' cyclic phosphate can only be ionized once. The

mobility change upon opening the cyclic phosphate is less than a factor of two (.53 to .82) because the cyclic phosphate has a smaller effective volume and is not retarded as much by the paper. For the same reason, these subspecies can be separated at pH 3.5. The R_F* =1.06 component migrates faster because it contains an additional negative charge caused by opening of the lactone ring to a δ-hydroxy acid (figure 9). This carboxylic acid is protonated at pH 3.5 and thus has only a minor effect on the mobility. The interconversion of R_F = .57 and R_F = .61 components at pH 3.5 is thus caused by the opening and closing of the lactone. This process is slow compared to the time scale of electrophoresis and hence the compounds can be separated because of their different effective volumes. The open and closed form of the lactone are in equilibrium at pH 3.5, but only the open, deprotonated form is observed at pH 8.8.

The opening of the lactone ring has been observed previously (Kanne et al., 1982) and occurs only in adducts at the 3,4 double bond. Reaction at this bond destroys the pseudo-aromaticity of the ring and facilitates opening. Thus, at pH 8.8, furan adducts have a charge of -2, coumarin adducts are -3, and crosslinks are -5. Cyclic phosphates complicate separation by lessening the charge so are removed by acid treatment prior to electrophoresis.

The UV spectra, ability to form crosslink, and mobility at pH 8.8 all indicate that the R_F = .57 component is a coumarin adduct and the R_F = .61 is a furan adduct. Short

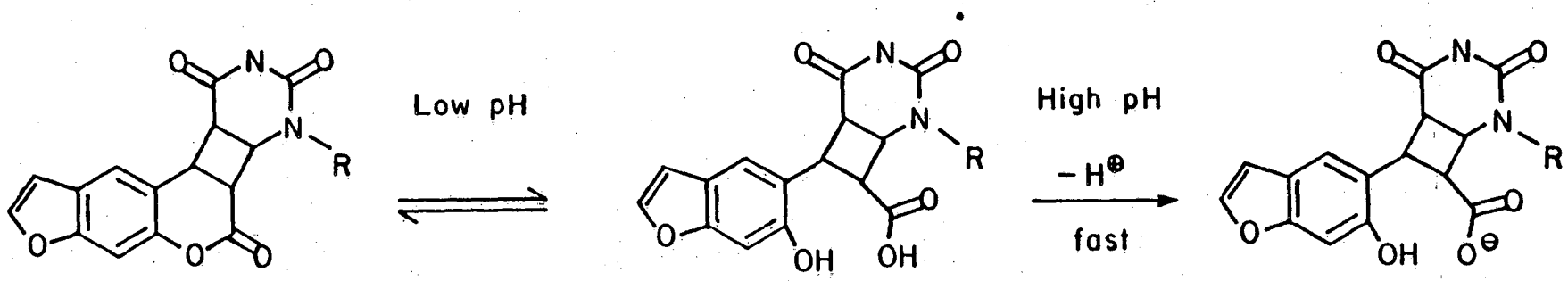

Low pH
High pH
$-H^{\oplus}$
fast
OH
OH
OH
$O^{\ominus}$

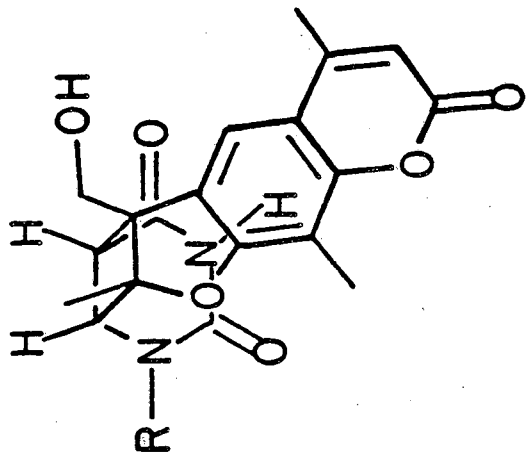
Figure 9: Equilibration of the coumarin adduct. At low pH, the coumarin adduct undergoes a reversible opening of the lactone ring. Both forward and reverse reactions are slow (on the order of hours) at room temerature. At sufficiently high pH, all of the adduct is driven to the open form because deprotanation of the acid prevents reclosure.

enzymatic digestions also lead to the transient formation of 2', 3' cyclic phosphates. This problem has arisen previously with unusual nucleosides, but it has not always been recognized (see, for example, Maden et al., 1975).

More rigorous assignment of structure has been performed on samples obtained by HPLC. NMR assignments were done by standard decoupling experiments and comparison to known structures. Complete details will be published at a later date.

D1 and D2 are diastereomers of the furan monoadduct to uridine. One of these adducts arises from reaction above (or on the 3' side) of the uridine while the other comes from reaction below (or on the 5' side) of the uridine ring (figure 10). Base adduct cannot be separated into two species because the chiral sugar is removed. The diastereomers thus become enantiomers which are only separable using chiral tools. The DNA analogs of these adducts have the same properties (Straub et al., 1981).

The equilibrium products have no equivalent in DNA adducts. Definitive structural analyses have not been possible because of the amount of material available. Certain properties have become evident. The entire coumarin ring structure must be intact because the furan adduct absorption spectrum is observed. The sugar cannot be a factor since the base adduct also undergoes the reaction. The equilibrium product is evidently more non-polar than the furan adduct because its mobility is less on the reverse phase column



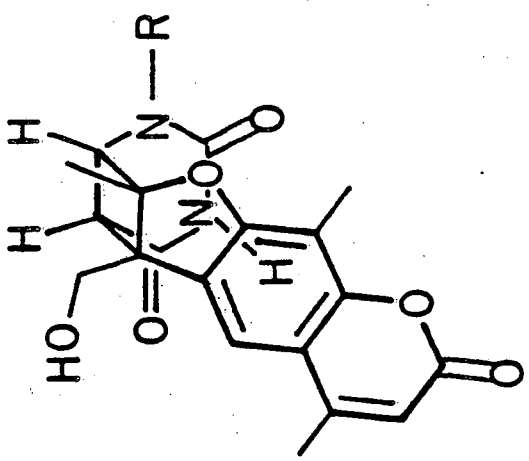


Figure 10: Stereochemistry of the furan diastereomer mono-adducts. The stereochemistry of the uridine adducts is shown, as determined by NMR and NOE enhancement. Nomenclature is from Straub et al. (1981).

and certain NMR resonances move to a more aliphatic region. The reaction probably will not be useful in devising chemical means of cleaving psoralen adducts because of its requirement for strong acid or very long reaction time.

Kinetics. The order of the reaction between UMP and HMT was determined by varying the concentration of UMP and determining the initial rate of reaction. If the amount of reacted material is very small compared to the total amount present, the following relation can be proven:

$$\nu_0 = k[X]^n$$

The order of reaction, n, can be found by simply varying the concentration of the reactant in question while keeping all other factors constant and then measuring the initial rate. Figure 6 demonstrates that the reaction is first order in UMP. The deviation from linearity at higher concentrations is caused by the breakdown of the above relationship because of the large amount of HMT reacted. It could also be caused by the presence of more than one type of mechanism for the reaction. The two lines shown in Figure 6 were obtained from a least squares fit of $\ln \nu_0 = n \ln[X] + C$. The slope of this line yields the best value of the exponent (order of reaction). For the four lowest concentrations only, a value of n = .99 was obtained. For all concentrations, a value of n = .90 fits the data best.

A first order reaction was to be expected based on the similarity of reaction rates of UMP and poly(U). If

the reaction was higher order, the high local concentrations of UMP in poly(U) would lead to increased reactivity. This rules out the possibility of an intercalation complex for the simple systems studied here.

The kinetics of formation of crosslink from monoadduct is shown in Figure 5. The tremendous excess of UMP and CMP used causes the reaction to be pseudo first order in these components. The standard method of graphing first order reactions is to plot the logarithm of the initial concentration of the starting material divided by the difference between the starting material and product versus time. The fact that straight lines are produced for both reactions indicates that they are first order in monoadduct.

The deviation from linearity observed is caused by the fact that not all of the starting material can be driven to crosslink. The monoadduct used in this experiment is derived from a mixture of E. coli tRNAs. Only 80% can form crosslinks in these conditions. This could be the result of the presence of unreactive isomers of the monoadduct (those with the 3,4 bond reacted) or photobreakdown products of the monoadduct which cannot react further but have the same mobility (e.g., photohydrates).

Relative Merits of Different Techniques. In previous studies of the reactions of psoralens with RNA and DNA, paper (Ou and Song, 1978) and thin layer chromatography (Pathak and Kramer, 1969) have been used to separate the photoproducts. These methods as well as changes in the

absorbtion or fluorescence have been used to monitor the progress of reaction (Lown and Sim, 1978; Dall'Acqua et al., 1979). Although these techniques have been very useful, we have found paper electrophoresis and HPLC to be more versatile in studying the complex products found in natural RNA as well as in monitoring the production of small amounts of products. HPLC gives much better resolution of products and allows better recovery of sample for further study than any other technique. Paper electrophoresis, while not giving as good resolution, can separate the major photoproducts and can be used for large numbers of samples more easily. These techniques are far more accurate and informative than techniques used previously.

The use of paper electrophoresis and HPLC has enabled us to examine the reaction of HMT and RNA in great detail. The ability to detect extremely small amounts of product (as low as 1 picogram under ideal circumstances) has made it possible to examine the reactions under a wide variety of conditions. Reactions with model compounds have allowed us to understand the kinetics and specificity of reaction. This knowledge has led to a better understanding of the way in which HMT reacts with natural RNAs <u>in vivo</u>.

References

Bachellerie, J. -P. and Hearst, J. E. (1982) Biochemistry 21, 1357-1363.

Dall' Acqua, F., Magno, S. M., Zambon, F., and Rodighiero, G. (1979) Photochem. Photobiol. 29, 489-495.

Hochkeppel, H. -K. and Gordon, J. (1979) Biochemistry 18, 2905-2910.

Isaacs, S. T., Hearst, J. E., and Rapoport, H. (1982) J. Labelled Compd. Radiopharm., in press.

Johnston, B. H., Johnson, M. A., Moore, C. B., and Hearst, J. E. (1977) Science 197, 906-908.

Kanne, D., Straub, K., Rapoport, H., and Hearst, J. E. (1982) Biochemistry 21, 861-871.

Krauch, C. H., Kramer, D. M., and Wacker, A. (1967) Photochem. Photobiol. 6, 341-354.

Lawley, P. D. and Brookes, P. (1963) Biochem. J. 89, 127.

Liu, F. J. and Yang, N. C. (1978) Biochemistry 17, 4865-4876.

Lown, J. W. and Sim, S. -K. (1978) Bioorg. Chem. 7, 85-95.

Maden, B. E. H., Forbes, J., deJonge, P. and Klootwijk, J. (1975) FEBS Lett. 59, 60-63.

Musajo, L., Bordin, F., Caporale, G., Marciani, S. and Rigatti, G. (1967) Photochem. Photobiol. 6, 711-719.

Ou, C. -N. and Song, P. S. (1978) Biochemistry 17, 1054-1059.

Pathak, M. A. and Kramer, D. M. (1969) Biochim. Biophys. Acta 195, 197-206.

Pathak, M. A., Kramer, D. M., and Fitzpatrick, T. B. (1974) in Sunlight and Man (ed. by Pathak et al.), pp. 335-368, Univ. of Tokyo Press, Tokyo.

Sanger, F., Brownlee, G. G., and Barrell, B. G. (1965) J. Mol. Biol. 13, 373-398.

Slae, S. and Shapiro, R. (1978) J. Org. Chem. 43, 4197-4200.

Sommer, S. S. (1979) Anal. Biochem. 98, 8-12.

Straub, K., Kanne, D., Hearst, J. E., and Rapoport, H. (1981) J. Am. Chem. Soc. 103, 2347-2355.

Townsend, L. B. and Robins, R. K. (1963) J. Am. Chem. Soc. 85, 242

Chapter Three:
Dependence of Psoralen Addition on the Conformation of RNA

INTRODUCTION

In order to take full advantage of psoralen's use as a structural probe, it is necessary to find both the specificity and the optimal conditions for the reaction. It has been apparent for some time that the most reactive base is uracil (Krauch et al., 1967; Pathak et al., 1974; Ou & Song, 1978; Bachellerie et al., 1981). Only more recently has it become known that adjacent uracils in a weak helix in natural RNA are a particularly strong site for addition (Bachellerie & Hearst, 1982; Thompson et al., 1981; Youvan & Hearst, 1982). In DNA, it has been suggested that an A-T site is preferred (Chandra et al., 1973; Dall'Acqua et al., 1978). Optimal conditions for psoralen reaction with 5S RNA (Thompson et al., 1981) and Col E1 DNA (Hyde & Hearst, 1978) have already been described. Studies with synthetic DNA polymers have determined relative reactivities but only in a narrow range of conditions (Dall'Acqua et al., 1979). Other studies have examined even more heterogeneous systems (Dall'Acqua et al., 1969).

We have chosen to study simpler systems in which a more homogeneous, controllable structure can be obtained. We have examined a wide range of salt concentrations, temper-

atures, and conformations for their effect on HMT incorporation. Circular dichroism and UV absorption spectroscopy have been used to correlate the structure of the polymers with HMT addition. This has been done with poly(U), poly(A,U)(random), poly(A-U)(alternating), and poly(A)+ poly(U). The synthetic polymers of adenine and uridine are worthwhile to study because they can adopt a variety of conformations and these have been well characterized by a variety of techniques (Stevens & Felsenfeld, 1964; Brahms, 1965; Massoulié, 1968; Steiner & Millar, 1970; Arnott & Bond, 1973). The transitions between different conformations are sharp and easily monitored. This work was performed with Jean Pierre Bachellerie and Kathi Hall, and appeared in Biochemistry vol. 20, pp. 1363-1368, March 1982.

MATERIALS AND METHODS

All polymers except poly(A-U) were purchased from P.L. Biochemicals. Poly(A-U) was synthesized by a modification of the method described by Chamberlin et al., (1963). The reaction (1.0 ml) was done with 50 mM Tris·HCl (pH 8.0), 1 mM $MnCl_2$, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 10 mM ATP, 10 mM UTP, 10 mM KCl, 10 μg/ml poly(dA-dT)(P-L Biochemicals), 1 μM 3'-O-methyl ATP, and 0.25 mg/ml E. coli RNA polymerase (a gift of Prof. Robert Woody and A. Young M. Woody of Colorado State University) for 2 hours at 37°. After reaction, the solution was phenol extracted three times, ethanol precipitated three times, dialyzed against 50 mM Tris pH 7.5

and 10 mM EDTA for 6 hours, followed by 6 hours with 1 mM Tris pH 7.5 and 0.1 mM EDTA. After ethanol precipitation, the sample was run through a Sephadex G100 column and the void fractions combined and precipitated. The yield was 800 μg. Aliquots of the reaction contained [α-^{32}P]UTP or [α ^{32}P]ATP (Amersham). Nearest neighbor analysis was done by digestion with T_2 RNase (Sigma) followed by paper electrophoresis at pH 3.5 (Bachellerie et al., 1981). Average length of the product was determined by electrophoresis on a 12% polyacrylamide gel (Thompson et al., 1981).

^{3}H-HMT with a specific activity of 4.0 x 10^7 cpm/μg was synthesized by S. Isaacs of this laboratory (Isaacs et al., 1982). In some experiments, 5'(^{32}P) labeled RNA was used. 1 mg of poly(U) was dephosphorylated with 1 unit of calf alkaline phosphatase (Boehringer-Mannheim) in 50 mM Tris pH 8.8 at 37° for one hour. The reaction mix was then phenol extracted twice, ether extracted, and ethanol precipitated. The polymer was redissolved in 50 mM Tris pH 8.5, 10mM $MgCl_2$, and 5 mM DTT. 50 μCi of [γ-^{32}P]ATP (Amersham, 3000 Ci/mmol) and 1 unit of T_4 polynucleotide kinase (Boehringer-Mannheim) were added and incubated for 30 min at 37°. The mixture was phenol extracted, ethanol precipitated, and run through a Sephadex G-100 column.

Unless otherwise noted, all irradiations were performed with 200 μg/ml of each polymer, 10 μg/ml ^{3}H-HMT, 1 mM Tris pH 7.5, 0.1 mM EDTA for 30 seconds or less. Irradiations involving poly(A,U) or poly(A-U) were done with 400 μg/ml

polymer in order to maintain the same amount of uridine in all cases. Samples containing both poly(A) and poly(U) were incubated at 50° for 1 hour and then slow cooled to the temperature of interest. Each incubation took place for a minimum of 18 hours. Irradiations were done as described in chapter 2.

After irradiation, samples were extracted twice with a 2x volume of chloroform, isoamyl alcohol (24:1, v/v) and then ethanol precipitated twice. The precipitate was resuspended and counted in a solution of toluene, triton X-100, and water (6:3:1, v/v/v) with 3.92 g/l PPO and 0.08 g/l Bis-MSB.

CD spectra were measured with a Cary 60 spectropolarimeter equipped with a 6003 unit. The computerized data collection system has been described previously (Brunner & Maestre, 1975). Data is presented as $\varepsilon_L - \varepsilon_R$ per mole of monomer, where ε_L and ε_R are the extinction coefficients for left and right circularly polarized light, respectively. Ultraviolet absorption spectra were measured with a Cary 14 spectrophotometer. In both instruments, the temperature of the sample was maintained with a circulating water bath. Melting curves for poly(A-U) were done with a Gilford Model 2527 thermoelectric temperature programmer. These were interfaced to a Commodore PET Model 2001 microcomputer.

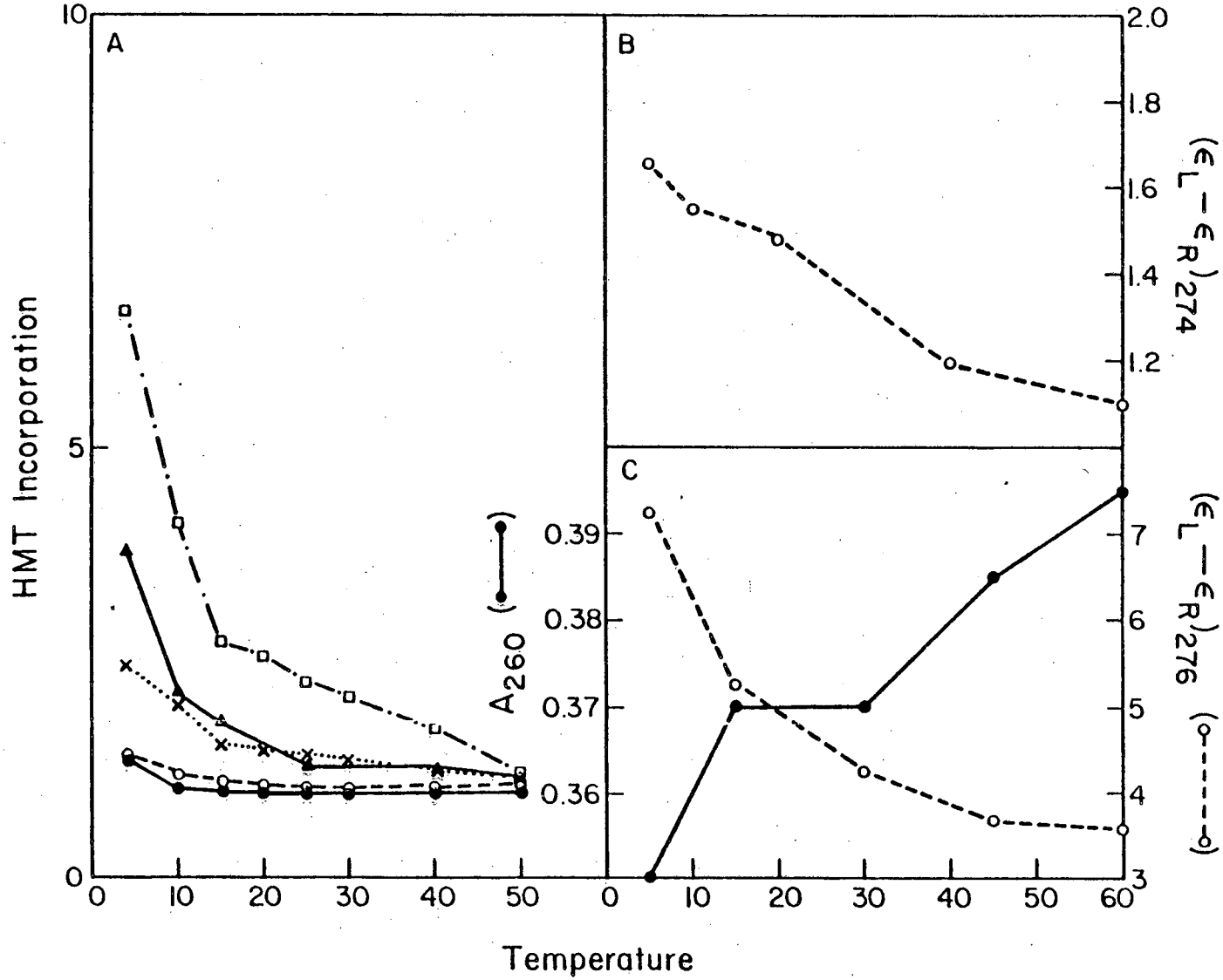
RESULTS

Poly U. The incorporation of ^{3}H-HMT into poly(U) as a function of temperature at different salt concentrations

is shown in Fig. 1a. The unit of HMT incorporation on the vertical axis is the same for Figs. 1, 2, 4, 5, 6, and 7 and is based on the reactivity of poly(U) in Tris-EDTA at 25° (equal to 1).

Random Poly(A,U). Figure 3 shows the incorporation of HMT into the random copolymer poly(A,U) in Tris-EDTA and in 0.1M NaCl. Figure 4 shows the temperature dependence of the CD spectra for these two conditions.

Alternating Poly(A-U). Before HMT incorporation studies were done on the alternating copolymer poly(A-U), its properties were determined in order to assure its authenticity. Nearest neighbor analysis using ∝-labeled triphosphates showed that the polymer was over 99% correct. The UV absorption spectra of both the coil and helix were identical to published spectra (Chamberlin et al., 1963) with isosbestic points at 279 nm and 295 nm. Carefully annealed polymer showed a hyperchromism of 67.2% at 260 nm, which is slightly larger than previously observed (65%). Fast cooled polymer showed a hyperchromism at 260 nm of 63.4%. These measurements were done with 5 mM NaCl between 20° and 80°. The melting points with 5 mM NaCl (42°) and 100 mM NaCl (65°) were sharp. Circular dichroism spectra of the helix and coil forms were identical to those previously published (Gray et al., 1972). The polymer was between 500 and 2000 bases long. The temperature dependence of HMT incorporation, UV absorption, and CD are shown in Fig. 4.


A
B
C
HMT Incorporation
A_{260} (
$(\epsilon_L - \epsilon_R)_{274}$
$(\epsilon_L - \epsilon_R)_{276}$ (o-----o)
Temperature
10
5
0
2.0
1.8
1.6
1.4
1.2
0.39
0.38
0.37
0.36
7
6
5
4
3
0 10 20 30 40 50
0 10 20 30 40 50 60

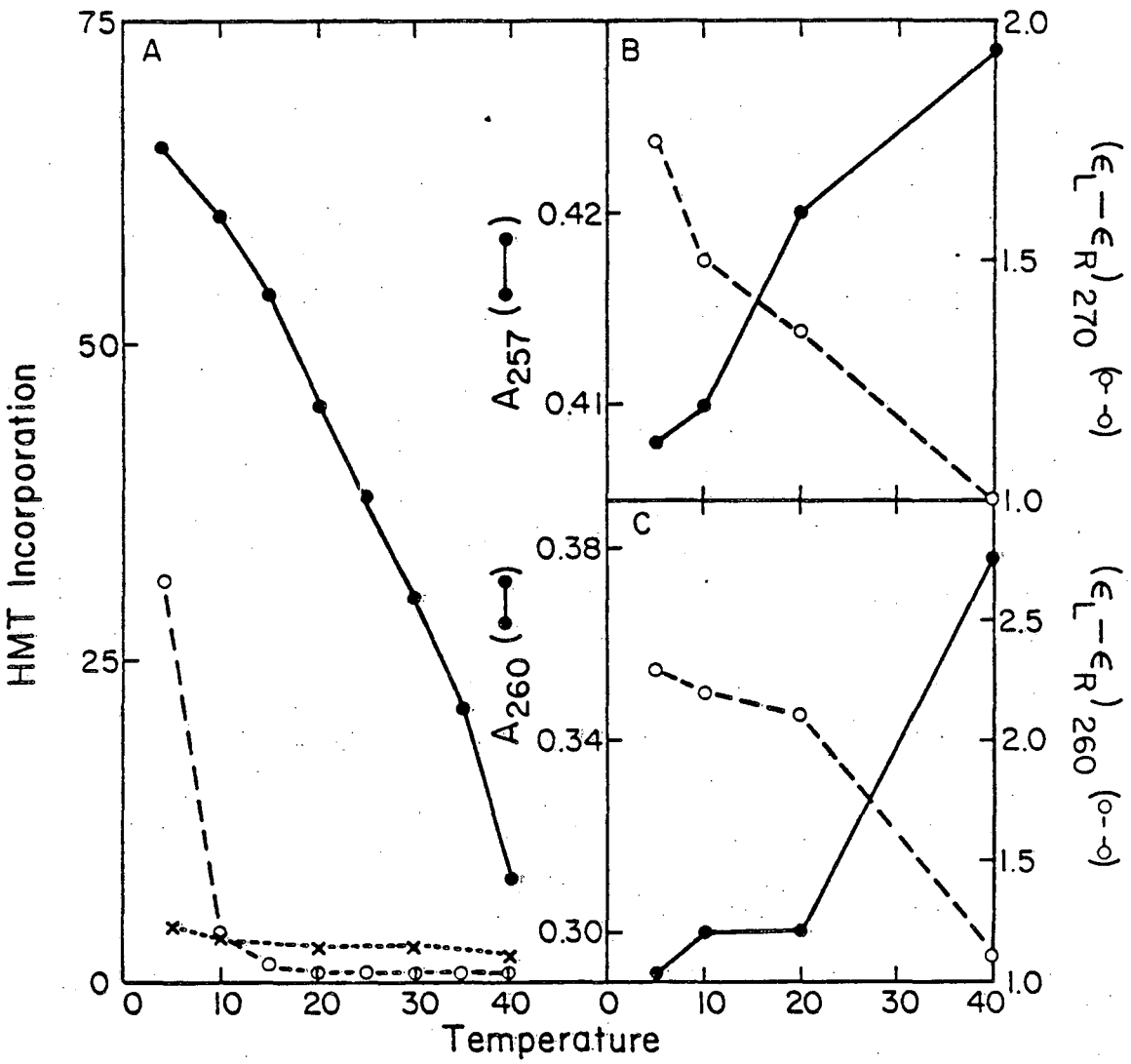
XBL 814-8996

Figure 1:

a) The temperature dependence of HMT addition to poly(U) at: no NaCl (•——•); 5 mM NaCl (•- - -•); 100 mM NaCl (×- - -×); 5 mM $MgCl_2$, 5 mM NaCl (▫- - -▫); 50 mM $MgCl_2$, 500 mM NaCl (▲- - -▲)

b) $(\varepsilon_L - \varepsilon_R)_{274}$ for poly(U) at 5 mM $MgCl_2$, 5 mM NaCl

c) $(\varepsilon_L - \varepsilon_R)_{276}$ for poly(U) at 50 mM $MgCl_2$, 500 mM NaCl (○- - -○) and absorption at 260 (•——•)


A
B
C
HMT Incorporation
Temperature
A_{257} (●—●)
A_{260} (●—●)
$(\epsilon_L - \epsilon_R)_{270}$ (o-o)
$(\epsilon_L - \epsilon_R)_{260}$ (o-o)
75
50
25
0
0
10
20
30
40
0.42
0.41
0.38
0.34
0.30
2.0
1.5
1.0
2.5

XBL 814-8995

Figure 2:

a) The temperature dependence of HMT addition to poly(A,U) at: no NaCl (o--o); 100 mM NaCl (•—•); and 5 mM NaCl, 5 mM $MgCl_2$ (×--×)

b) $(\epsilon_L-\epsilon_R)_{270}$ for poly(A,U) with no NaCl (o--o) and absorption at 260 nm (•—•)

c) $(\epsilon_L-\epsilon_R)_{260}$ for poly(A,U) at 100 mM NaCl (o--o) and absorption at 260 nm (•—•)

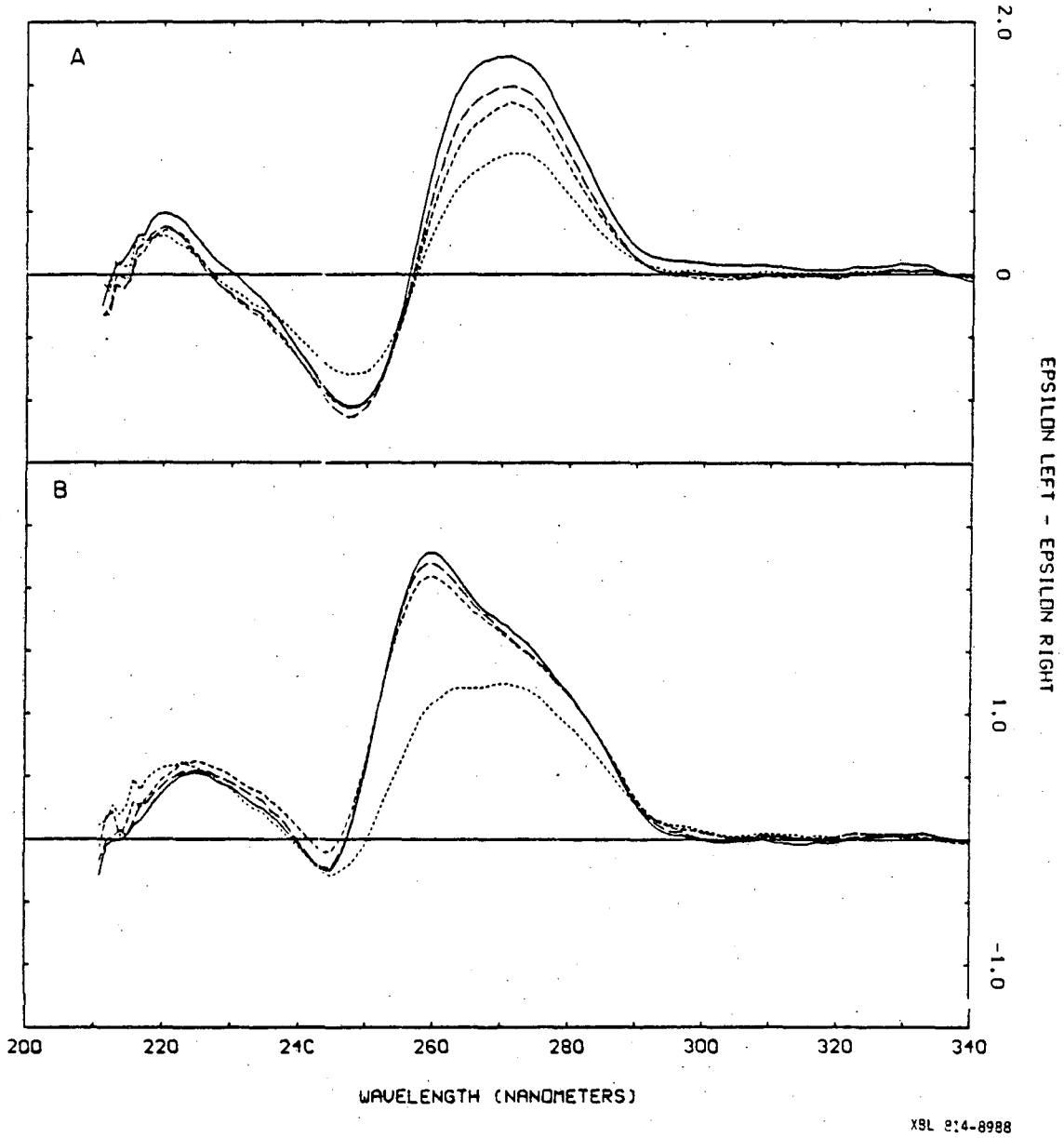

A
B
2.0
0
1.0
-1.0
EPSILON LEFT - EPSILON RIGHT
200
220
240
260
280
300
320
340
WAVELENGTH (NANOMETERS)
XBL 814-8988

Figure 3:

a) The circular dichroism spectrum of poly(A,U) with no NaCl at 5° (——), 10° (— — —), 20° (- - - -), 40° (· · · · ·).

b) The circular dichroism spectrum of poly(A,U) with 100 mM NaCl at 5° (——), 10° (— —), 20° (- - - -), 40° (· · · · ·).

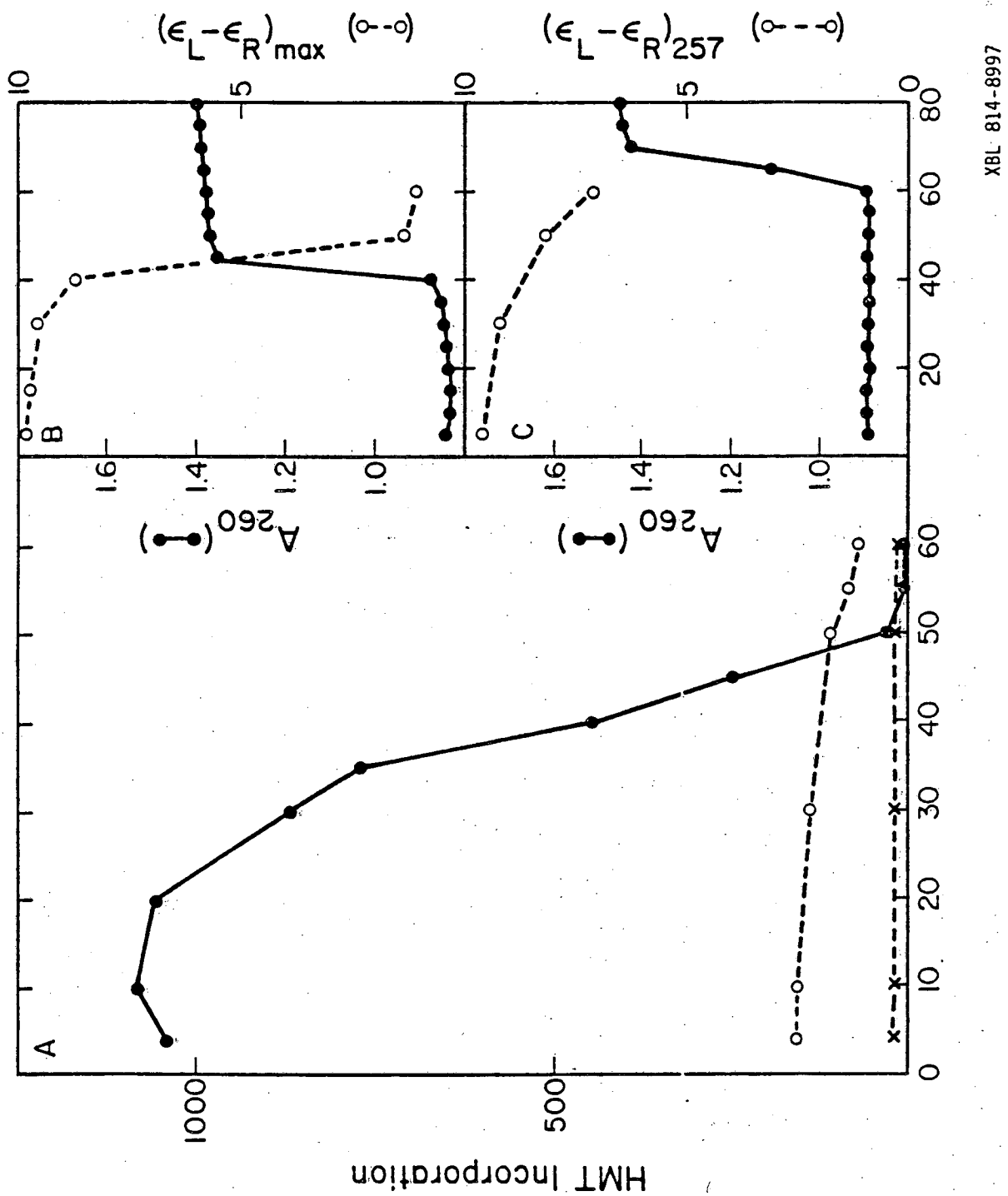

$(\epsilon_L-\epsilon_R)_{max}$ (o--o)
$(\epsilon_L-\epsilon_R)_{257}$ (o--o)
B
C
A
A_{260} (●—●)
A_{260} (●—●)
HMT Incorporation
XBL 814-8997

Figure 4:

a) The temperature dependence of HMT addition to poly(A-U) at: 5 mM NaCl (●——●), 100 mM NaCl (○- -○), and 5 mM $MgCl_2$, 5 mM NaCl (×- -×)

b) $(\epsilon_L - \epsilon_R)_{max}$ for poly(A-U) at 5 mM NaCl (○- -○) and absorption at 260 nm (●——●)

c) $(\epsilon_L - \epsilon_R)_{257}$ for poly(A-U) at 100 mM NaCl (○- -○) and absorption at 260 nm (●——●)

Double-Stranded Poly(A)·Poly(U) and Triple-Stranded Poly(A)·2Poly(U). By proper choice of salt condition and molar ratio of homopolymers, it is possible to form either a double or triple helical structure (Stevens & Felsenfeld, 1964). The HMT reactivity of double-stranded poly(U)·poly(A) was examined in a wide range of temperature and salt concentrations. After heating and slow cooling to the temperature of interest, an equimolar mixture of poly(U) and poly(A) was photoreacted with HMT at different salt concentrations and temperatures (figure 5).

The rate of HMT incorporation into poly(A)·poly(U) was examined over a wide range of salt concentrations as shown in Fig. 6. At 4°, the amount of incorporation fell logarithmically between 5 and 100 mM NaCl and then tailed off more slowly. The salt dependence at 37° was monitored at two poly(A) to poly(U) ratios, 1:1 and 1:2. The equimolar mixture curve is virtually identical to that at 4° except that incorporation does not increase until 10 mM NaCl (the lowest salt at which the double-stranded complex is stable). The salt dependence of the mixture which contains only half as much poly(A) has the same shape at low salt but drops off more quickly at higher salt.

In order to determine the rate of HMT addition to the triple-stranded poly(A)·2poly(U), the amount of poly(A) was varied while keeping poly(U) constant. This is shown in Fig. 7. In conditions where the triple-stranded complex cannot form (25°, 10 mM NaCl) (Stevens & Felsen-

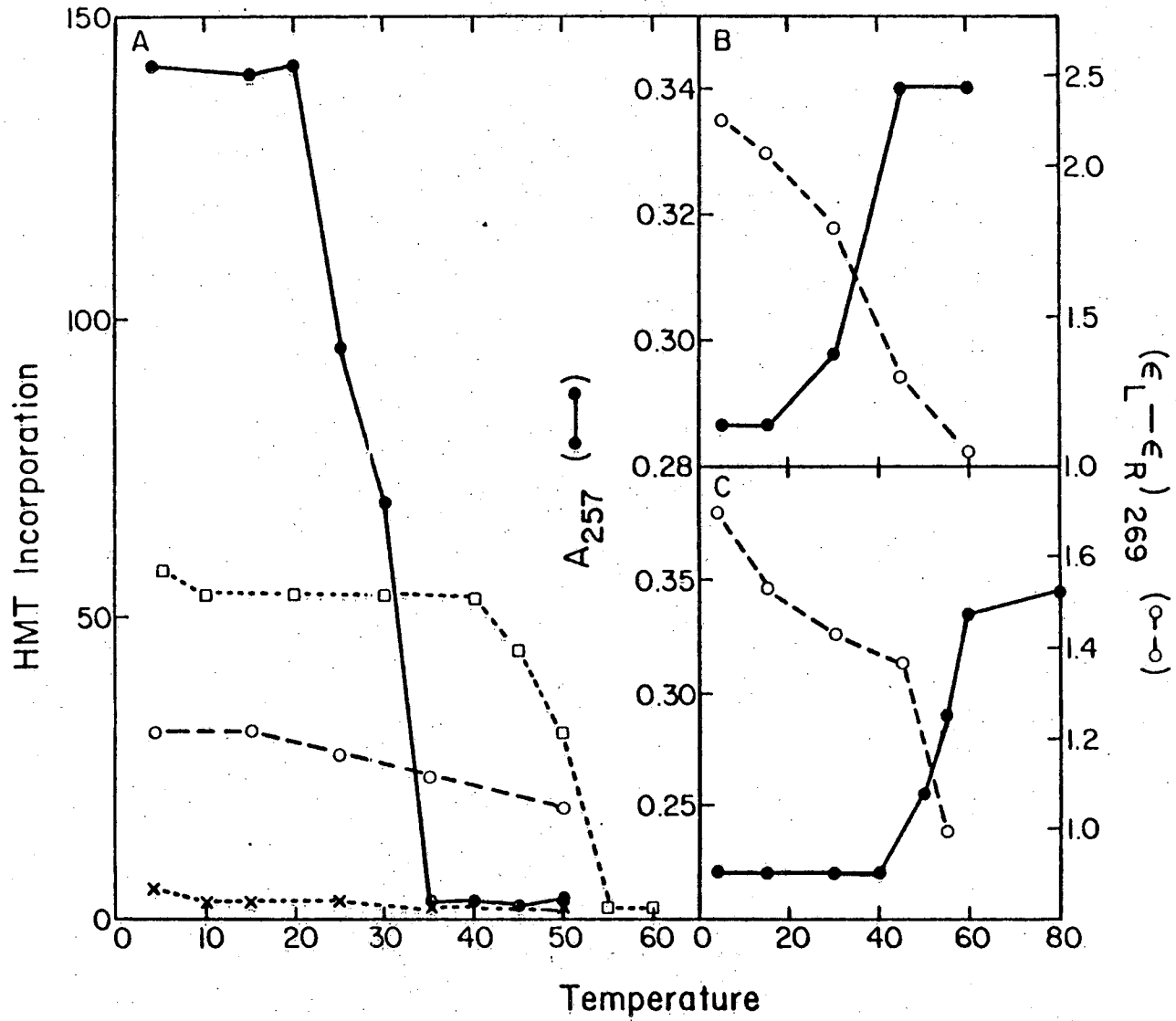

A
B
C
HMT Incorporation
Temperature
A_{257} (●—●)
$(\epsilon_L - \epsilon_R)_{269}$ (○—○)
150
100
50
0
0.34
0.32
0.30
0.28
0.35
0.30
0.25
2.5
2.0
1.5
1.0
1.6
1.4
1.2
1.0
0 10 20 30 40 50 60
0 20 40 60 80

Figure 5:

a) The temperature dependence of HMT addition to poly(A)·poly(U) (equal mole amounts of each polymer) at: 5 mM NaCl (•——•); 50 mM NaCl (•——•); 100 mM NaCl (•- -•); 5 mM $MgCl_2$, 5 mM NaCl (•- -•)

b) $(\epsilon_L - \epsilon_R)_{269}$ for poly(A)·poly(U) at 5 mM NaCl (•- -•) and absorption at 257 nm (•——•)

c) $(\epsilon_L - \epsilon_R)_{270}$ for poly(A)·poly(U) at 50 mM NaCl (•- -•) and absorption at 257 nm (•——•)

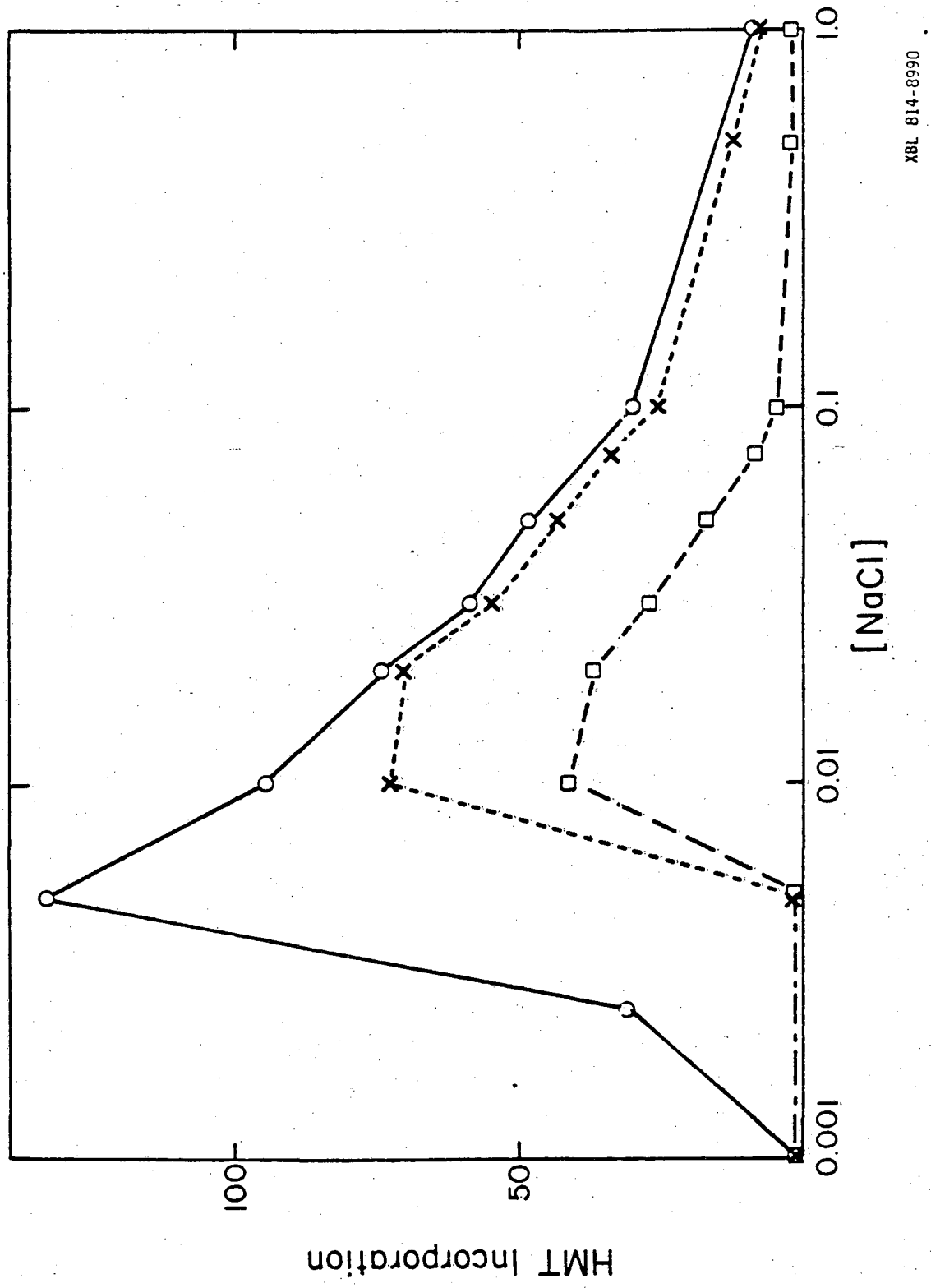

HMT Incorporation
100
50
0.001
0.01
0.1
1.0
[NaCl]
XBL 814-8990

Figure 6:

The concentration of NaCl was varied for poly(A) (200 µg/ml) and poly(U) (200 µg/ml) at 4° (●—●) and 37° (×···×) and for poly(A) (100 µg/ml) and poly(U) (200 µg/ml) at 37° (□—□).

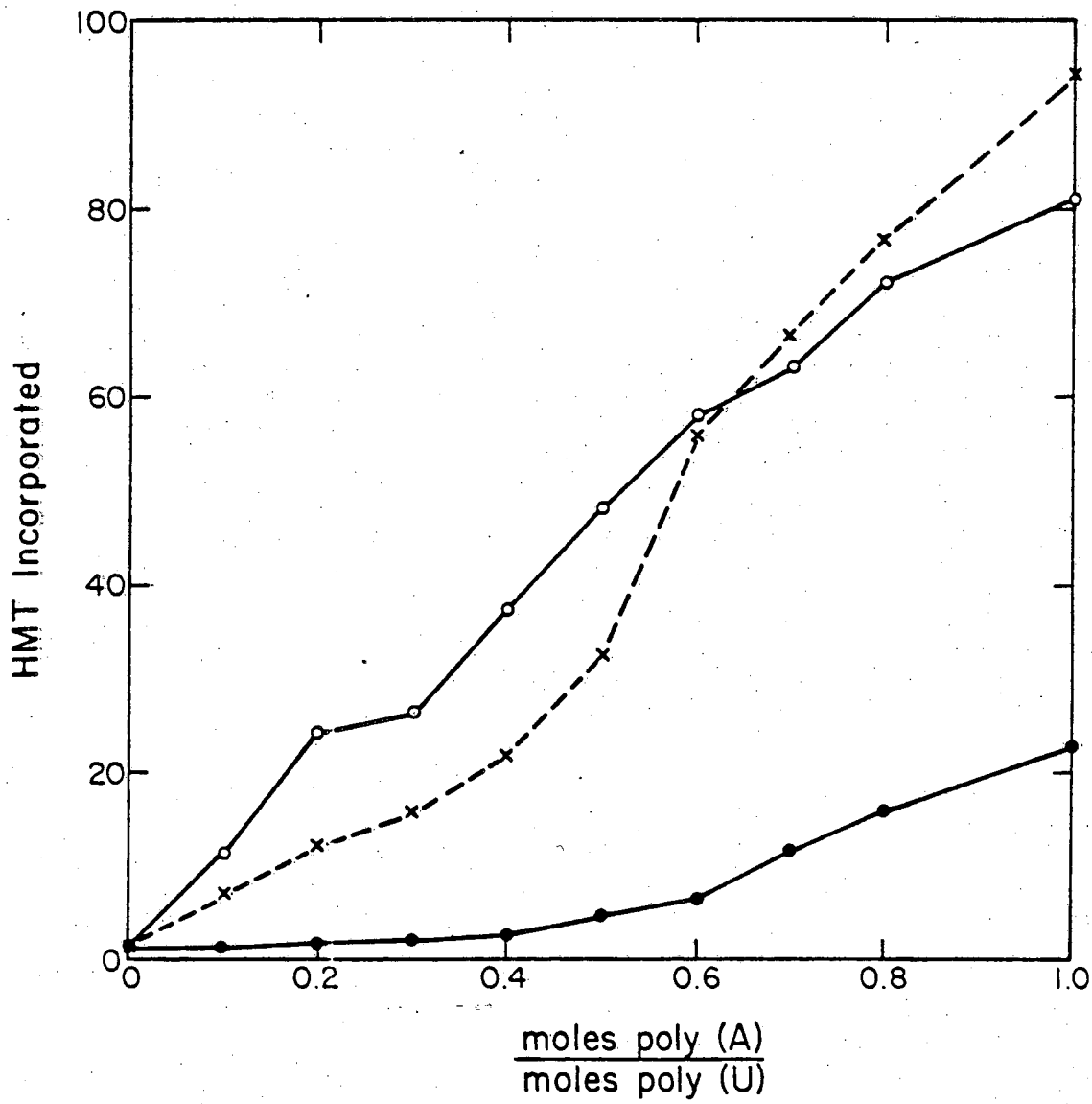

100
80
60
40
20
0
HMT Incorporated
0
0.2
0.4
0.6
0.8
1.0
moles poly (A)
moles poly (U)

Figure 7:

Poly(U) (200.μg/ml) was mixed with varying amounts of poly (A) at: 25°, 10 mM NaCl (○——○); 4°, 10 mM NaCl (×- - -×); 4° and 25°, 100 mM NaCl (●——●).

feld, 1964; Massoulié, 1968), a linear increase in the rate of HMT addition with added poly(A) is seen. The amount of HMT incorporated reaches a plateau at a molar ratio of slightly more than 1:1 (A:U). In conditions where both the triple and double-stranded complexes can exist (25°, 100 mM NaCl) the amount of HMT incorporated increases only slightly with increasing poly(A) until a ratio of 1:2 is reached. At this point, the amount of HMT added increases quickly until an A:U ratio of 1:1 is reached. This experiment was repeated with the same salt conditions but at 4°. At 100 mM NaCl, the incorporation profiles are identical at the two temperatures. At 10 mM NaCl, the rate of addition is lower than would be predicted for a double-stranded complex below a 1:2 ratio of A:U. This effect is not nearly as dramatic as at 100 mM NaCl.

DISCUSSION

The conditions used to photoreact HMT with RNA in this study were chosen to be as gentle as possible. The irradiation was short in order to prevent crosslinking. Quantitation of crosslinks by paper electrophoresis of hydrolysates (Bachellerie et al., 1981) showed there was a negligible amount. The binding constant of HMT is sufficiently low (Isaacs et al., 1977) that the noncovalently bound HMT has a minimal impact on the polymer structure. This is not true for intercalators with a higher binding constant such as ethidium bromide. Ethidium bromide can unfold the tertiary structure of tRNA but

stabilizes other polymers (Urbanke et al., 1973). It has also been shown to disproportionate triple-stranded 2poly(U)·poly(A) at high concentrations (Bresloff and Crothers, 1981). HMT does not have as large an effect because not as many molecules are bound.

Ultraviolet absorption and circular dichroism are useful probes for studying the structures of our RNA systems because they monitor different aspects of the geometry. UV absorption measures mainly the stacking between bases and hence is a good method of observing secondary structure, but it is relatively insensitive to higher order structure. CD, on the other hand, arises principally from the interaction between optical transitions of the bases. The theory for predicting the structural changes which are occurring when the CD changes is not well advanced yet, but progress is being made (Greve et al., 1978).

Poly(U). In low salt, poly(U) has no secondary structure. At 0.1 M NaCl, poly(U) is optically identical to the polymer at low salt but the level of HMT addition is slightly elevated, particularly at low temperature. This could be caused by a small amount of structure which is not discernible by CD.

When 5 mM $MgCl_2$ is present, CD indicates that poly(U) has secondary structure. This is not apparent from the UV absorption, however. The latter is a result of the poor base stacking properties of uridine. Thrierr et al.

(1971) have shown that poly(U) folds back on itself at high salt to form a double-stranded structure. Even though the base pairing has been shown to be much different than the classical Watson-Crick type (Young & Kallenbach, 1978), HMT can intercalate and react with the uridine. The CD spectrum suggests that the helix is an A-form, similar to normal double-stranded RNA. It is not clear whether the intercalation site in the poly(U) is essentially the same as in other polymer structures. At very high salt (500 mM NaCl, 50 mM $MgCl_2$), both absorption and CD spectra indicate a stable structure exists. In fact, the structure in the poly(U) is stabilized so much that intercalation is made difficult. Because intercalation is accompanied by an unwinding of the helix (Wiesehahn & Hearst, 1978) and a lengthening of the RNA, anything which stabilizes the helix and condenses the RNA will inhibit intercalation. Although the rate of HMT addition falls at these high salt concentrations, it is still higher than in the randomly coiled poly(U).

Random Poly(A,U). As determined by CD, the structure of the random copolymer poly(A,U) is substantially different in no NaCl and 100 mM NaCl. In Tris-EDTA, the CD is characteristic of a B-form structure with positive and then a negative band of approximately equal strength (a conservative spectrum) (Tinoco et al., 1980). This type of CD is very uncharacteristic of double-stranded RNA. Since the ionic strength is low and would not shield the

phosphate charges effectively upon double strand formation, it is uncertain that any base pairing is present. The spectrum is more like that observed with adenylate oligomers (Brahms et al., 1966). As observed with these oligomers, the peaks decrease in intensity and are red-shifted with increasing temperature. Since short runs of adenines are present in the polymer, it seems likely that this type of interaction could be occurring. Neither the CD nor the UV absorption predict the temperature dependence of the HMT incorporation. It is possible that, even in the absence of base pairing, the stacking of adenines is sufficient to create local order in the uridine residues at very low temperatures. This ordered structure would allow intercalation and reaction. Because uridine does not stack well, its contribution to the change in the CD and UV absorption would be small.

The CD of poly(A,U) in 0.1 M NaCl is more complex. It is much more like the normal A-form RNA spectrum with a large positive band and a small negative band (nonconservative spectrum). There are clearly at least two components in the positive band. It seems likely that normal Watson-Crick base-pairing is present, although the type of structure postulated for the low salt poly(A,U) could also be present. The large change in optical properties occurring between 20° and 40° is probably caused by a loss of base-pairing. The drop in HMT incorporation is not as sharp, but falls almost

linearly between 5° and 40°.

Alternating Poly(A-U). Although the shape of the CD for poly(A-U) suggests an A-form helix, there are differences between it and the normal RNA spectrum (Gray et al., 1972). Below the melting temperature, UV absorption is essentially constant while HMT incorporation and CD decrease somewhat with increasing temperature. Of the synthetic polymers studied, poly(A-U) has the highest rate of HMT reaction, with over 1000 times more HMT incorporated than poly(U). In DNA, it was also observed that the alternating purine-pyrimidine polymers had higher reactivities than the homopolymers (Dall'Acqua et al., 1979). This does not appear to be caused by the rate of dark binding. Binding of ethidium bromide to homopolymers is much higher than to alternating purine-pyrimidine polymers. (Bresloff and Crothers, 1981). Apparently, the alignment of photoreactive bonds is better in alternating structures.

Double-Stranded Poly(A)·Poly(U) and Triple-Stranded Poly(A)·2Poly(U). When poly(A) and poly(U) are mixed in equal proportions in the range 5-100 mM NaCl, the CD spectrum indicates that the polymers adopt an A-form geometry. The CD spectra are more complex than normally associated with this geometry, but contain many A-form features. In both 5 and 50 mM NaCl, the HMT incorporation falls off at temperatures slightly below the optical melting temperatures.

Poly(A) and poly(U) are also capable of forming triple helices. Depending on the salt, temperature, and molar ratios, single-, double-, or triple-stranded complexes may be present in solution. We have examined conditions in which the double-stranded complex is the most stable over all molar ratios of A:U and conditions in which either the double- or triple-stranded complex can form, depending on the ratio in which the polymers are mixed. Although conditions exist in which only the triple helix is stable, very high salt and temperature are necessary. This produces very low levels of HMT incorporation and thus changes are not easily measured.

At 10 mM NaCl and 25°, only the double-stranded complex is stable. As expected, HMT incorporation increases linearly with poly(A) added (Fig. 7). Once all of the poly(U) has been base paired to poly(A), the incorporation levels off. In conditions where both double- and triple-stranded complexes can form (100 mM NaCl at 4° and 25°, 10 mM NaCl at 4°), HMT incorporation is not a simple function of poly(A) added. In all cases, the amount of HMT reacted is substantially lower with the triple helix than would have been predicted for a double helix. When sufficient poly(A) has been added to start formation of double-stranded molecules, the level of HMT incorporation rises dramatically.

At 10 mM NaCl and 4°, the incorporation curve can be approximated by two straight lines. If the line

connecting the points up to a poly(A)/poly(U) molar ratio of 0.5 is extrapolated, it predicts that the plateau value for HMT incorporation would be one half of what it actually turns out to be. One interpretation of this is that the strand of poly(U) which is involved in Watson-Crick base pairing reacts normally while the strand which is in the major groove running parallel to the poly(A) (Arnott & Bond, 1973) does not react to an appreciable extent. At 100 mM NaCl, the third strand could be so tightly bound that it inhibits reaction of both strands of poly(U). Alternatively, the structure of the strands which are involved in Watson-Crick base-pairing might be perturbed sufficiently that the photoreactive bonds are not aligned correctly after intercalation occurs.

Evidence for reduced incorporation after triple-strand formation may also be seen in Figure 6. At 37° and a 1:1 ratio of polymers, double-stranded complexes form at all salt concentrations above 10 mM. At a 1:2 (A:U) ratio, the situation is more complex. At 1 and 5 mM NaCl, only single strands are present. Only single- and double-stranded molecules should be present from 10 to 30 mM NaCl. As expected, the lower amount of poly(A) leads to reduced incorporation of HMT (50-60% of the 1:1 ratio). Above 50 mM NaCl, both double and triple strands should be present. The amount of HMT incorporation relative to the 1:1 ratio drops. At 50 mM NaCl, the level of incorporation is 40% of the 1:1 mixture, 25% at

75 mM NaCl, and less than 15% at 100 mM NaCl and above.

The fact that the triple-stranded poly(A)•2poly(U) is much less reactive to HMT than the double-stranded complex is not surprising. X-ray diffraction studies have shown that the triple helix has virtually the same radius and pitch as the double helix (Arnott & Bond, 1973). It is thus much more condensed and difficult, if not impossible, to unwind. It seems probable that a marginally stable triple helix does allow some intercalation, possibly with the third strand looping out. Another possibility is that imperfections in the helix arise which allow double-stranded regions to exist.

Insight into how RNA conformation affects reactivity can also be gained by examining the types of HMT adducts formed as well as the number. This has been done for a number of polymers, as shown in table I, using paper electrophoresis at pH 8.8 for separation. The same experiments, using HPLC for separation, have been done for a smaller set of polymers. It should be noted that the numbers obtained by the two methods are in qualitative, but not quantitative, agreement. One reason for this discrepancy is the fact that samples tested with both methods have not been prepared in identical fashions. Also, and perhaps more importantly, is the problem of decomposition of the coumarin adduct. As mentioned in chapter 2, this adduct opens up to form an acid. This is very unstable and makes HPLC analysis particularly

difficult because the charge grossly affects the mobility. The two methods can be reconciled if the material which is unaccounted for (in the void volume or unrecoverable) is attributed principally to coumarin adduct. A more careful, complete analysis should resolve this problem.

Despite the lack of quantitative agreement between methods, paper electrophoresis is reproducible and can yield important information. The fraction of coumarin adduct which cannot be driven to crosslink varies over a factor of ten. The only completely unstructured polymer, poly(U), has the largest amount. Uridine and UMP have even larger amounts of coumarin adduct. This indicates that, as predicted by quantum mechanical calculations (Song et al., 1971), the 3,4 bond is inherently more reactive than the 4',5' bond. As the polymers become more structured and more reactive, there is a general decrease in the fraction of coumarin adduct. This indicates that the geometry of the intercalation complex is the principal driving force in determining reactivity. The effect of the NaCl concentration on the fraction of coumarin adduct is small (when comparing the same polymer).

The time dependence of coumarin adduct formation indicates there is more than one class of reactive site. With poly(A,U) and poly(A-U) at both 5 and 100 mM NaCl, there is a marked increase in coumarin adduct with time. With poly(A)+ poly(U), there is a slight decrease in coumarin adduct at longer times. If all reactive sites

Table I: Conformation dependence of photoadduct production

Polymer	[NaCl] (mM)	Furan Monoadduct		Coumarin Monoadduct		Crosslink	
		0.5 min	10 min	0.5 min	10 min	0.5 min	10 min
Poly(U)	0	-	44.5	-	30.6	-	24.9
Poly(U) + 500 mM $MgCl_2$	50	-	32.5	-	22.7	-	44.3
Poly(A,U)	0	99.1	65.5	0.9	11.0	0	23.5
Poly(A,U)	100	95.3	54.6	4.0	12.9	0.7	32.5
Poly(A)·Poly(U)	5	80.1	78.7	19.9	19.0	0	2.3
Poly(A)·Poly(U) @ 20°	5	-	77.5	-	21.3	-	1.1
Poly(A)·Poly(U)	100	78.9	79.8	21.1	17.6	0	2.6
Poly(A)·2Poly(U)	100	84.8	70.4	15.2	12.5	0	17.1
Poly(A-U)	5	99.6	88.4	0.4	3.1	0	10.0
Poly(A-U) @ 20°	5	-	94.8	-	0.9	-	4.2
5S RNA	5	-	70.1	-	9.1	-	20.8
5S RNA @ 20°	5	-	81.0	-	6.4	-	12.6

Irradiations were performed for the time indicated at 5°, unless otherwise stated.

were identical, there should be no change in the fraction of coumarin adduct. As expected, there is very little crosslink formed at short times.

The amount of crosslink formed is also highly dependent on conformation. Perhaps the most surprising result is the small amount of crosslink in poly(A-U), where every furan adduct is a potential crosslink. In the analogous DNA polymer, using the same conditions, more than 90% of the furan adduct is converted to crosslink (Kanne et al., 1982). This clearly demonstrates that reactivity is not simply determined by the availibility of a site but also requires the correct conformation. Equally surprising is the large amount of crosslinking in random poly(U). This aptly demonstrates the dynamics of the polymer in solution, as remote parts of the molecule must come in close contact with the HMT adducts to permit crosslinking. The amount of crosslinking in poly(A)·poly(U), while very small, indicates there must be imperfections in the helix or breathing motions which allow two uridines to crosslink.

While the temperature and salt effects on the amount of coumarin adduct formed are small, the effects on crosslink formation are larger and more uniform. In the two cases where comparison can be properly made (poly(A)·poly(U) and poly(A-U)), increasing of NaCl from 5 to 100 mM increases crosslinking by about 10%. Temperature has a more profound effect. Increasing the temperature from 4° to 20° decreases crosslinking by 39 to 68%. Clearly,

crosslinking is not dependent solely on thermal breathing to allow the correct conformation (which would be higher at higher temperatures). The low temperature structure must simply be a better structure for crosslinking.

HPLC allows an even more detailed analysis of products to be made, but these experiments have not been done systematically. It is noteworthy that the amount of each of the two diastereomers which make up the furan adduct varies considerably in the three polymers tested. Yeast tRNA yields equal amounts of the two while poly(A)·poly(U) in 5 mM NaCl has a D2:D1 ratio of 10:1 and poly(A,U) at 100 mM NaCl has a ratio of 1:2. The significance of this is not yet clear but it may eventually be possible to glean structural information from this type of data.

In the tremendous variety of conditions used in these experiments, we never found the amount of HMT incorporated into poly(U) to be lower than that found in the random coil. This implies two types of reaction. The preferred mode of reaction is preceded by intercalation (Dall'Acqua et al., 1978a; Tjerneld et al., 1979). This allows π bond stacking which stabilizes the structure such that the reactive bonds of the HMT and uridine are optimally situated for photoreaction. This can lead to a high rate of incorporation as seen in poly(A-U) at low salt and temperature. Clearly, this type of interaction cannot take place when poly(U) is in a random coil or a very tightly packed helix. It is unlikely that any type of overlap between the base and the HMT could

occur in a tight helix, so the low level of reactivity must arise from either random encounters from the solution or some type of weak, outside binding. HPLC analysis of the stereoisomers of the monoadducts produced with each compound should prove useful in determining the type of interactions taking place.

From this study, it is obvious that certain types of helical structure accelerate the reaction of HMT with RNA. In general, low salt and low temperature favor HMT reaction. These conditions lead to a relatively weak helix which can be easily unwound. When salt is added, it becomes difficult for the drug to intercalate between the more tightly packed base pairs. Magnesium has a particularly strong inhibitory effect. The fact that HMT reactivity is intimately linked to RNA structure is further exemplified by the same type of logarithmic dependence on salt concentration as is observed for melting temperature (Dove and Davidson, 1962). At higher temperatures, the thermal fluctuations in the RNA shorten the lifetime of the intercalation complex and thus lower the reaction rate. Temperature has a greater effect on the dissociation constant than on the association constant (Hyde and Hearst, 1978). In more complex systems, these generalities do not always hold true. For tRNA, an increase in HMT reactivity with increasing temperature was observed at one salt concentration because of unfolding of the tertiary structure (Bachellerie and Hearst, 1982).

Changes in the rate of HMT incorporation could not be

perfectly predicted by either CD or UV absorption transitions. These three probes of RNA conformation are influenced by different aspects of structure, so it is not surprising that they yield slightly different results. It is not trivial to extrapolate the structure-reactivity relationships found in synthetic polymers to the more complex RNAs of biological interest. Also, there are many structures found in natural RNAs that cannot be duplicated with synthetic polymers. Despite these complications, study of these model compounds does provide insight into what types of factors are likely to influence HMT reactivity. Combining the knowledge of what conformations are most reactive to HMT with the known base specificity (chapters 2 and 6) should allow workers to choose systems and conditions which are most likely to yield worthwhile results.

REFERENCES

Arnott, S., & Bond, P. J. (1973) Nature New Biol.244, 99-101.

Bachellerie, J.-P., & Hearst, J. E. (1982) Biochemistry 21, 1357-1363.

Bachellerie, J.-P., Thompson, J. F., Wegnez, M. R., & Hearst, J. E., (1981) Nucl. Acids Res. 9, 2207-2222.

Brahms, J. (1965) J. Mol. Biol. 11, 785-801.

Brahms, J., Michelson, A. M., & Van Holde, K. E. (1966) J. Mol. Biol. 15, 467-488.

Bresloff, J. L. & Crothers, D. M. (1981) Biochemistry 20 3547-3553.

Brunner, W. C., & Maestre, M. F. (1975) Biopolymers 14, 555-565.

Chamberlin, M. J., Baldwin, R. L., & Berg, P. (1963) J. Mol. Biol. 7, 334-349.

Chandra, P., Mardiani, S., Dall'Acqua, F., Vedaldi, D., Rodighiero, G., & Biswa, R. K. (1973) FEBS Lett. 35, 243-246.

Dall'Acqua, F., Marciani, S., & Rodighiero, G. (1969) Z. Naturforsch 24B, 307-314.

Dall'Acqua, F., Terbojevich, M., Marciani, S., Vedaldi, D., & Recher, M. (1978) Chem.-Biol. Interactions 21, 103-115.

Dall'Acqua, F., Vedaldi, D., & Recher, M. (1978) Photochem. Photobiol. 27, 33-36.

Dall'Acqua, F., Vedaldi, D., Bordin, F., & Rodighiero, G. (1979) J. Invest. Derm. 73, 191-197.

Dove, W. F. & Davidson, N. (1962) J. Mol. Biol. 5, 467-478.

Gray, D. M., Tinoco, Jr., I., & Chamberlin, M. J. (1972) Biopolymers 11, 1235-1258.

Greve, J. Maestre, M. F., Moise, H., & Hosoda, J. (1978) Biochemistry 17, 887-893.

Hyde, J. E. & Hearst, J. E. (1978) Biochemistry 17, 1251-1257.

Isaacs, S. T., Hearst, J. E., & Rapoport, H. (1982) J. Labeled Compd. Radiopharm., in press.

Isaacs, S. T., Shen, C. -K. J., Hearst, J. E., & Rapoport, H. (1977) Biochemistry 16, 1058-1064.

Kanne, D., Straub, K., Hearst, J. E., & Rapoport, H. (1982) J. Am. Chem. Soc., in press.

Krauch, C. H., Kramer, D. M., & Wacker, A. (1967) Photochem. Photobiol. 6, 341-354.

Massoulie, J. (1968) Eur. J. Biochem. 3, 428-438.

Ou, C. -N. & Song, P. -S. (1978) Biochemistry 17, 1054-1059.

Pathak, M. A., Kramer, D. M., & Fitzpatrick, T. B. (1974) in Sunlight and Man (Pathak et al., eds.) pp. 335-368, University of Tokyo Press, Tokyo.

Rabin, D. & Crothers, D. M. (1979) Nucl. Acids Res. 7, 689-703.

Song, P. -S., Harter, M. L., Moore, T. A., and Herndon, W. C. (1971) Photochem. Photobiol. 14, 521-530.

Steiner, R. & Millar, D. B. S. (1970) in Biological Polyelectrolytes, Vol. 3 (Veis, A., ed.) pp. 65-120, Marcel Dekker, Inc., New York.

Stevens, C. L. & Felsenfeld, G. (1964) Biopolymers 2, 293-314.

Thompson, J. F., Wegnez, M. R., & Hearst, J. E. (1981) J. Mol. Biol. 147, 417-436.

Thrierr, J. C., Dourlent, M., & Leng, M. (1971) J. Mol. Biol. 58, 815-830.

Tinoco, I. Jr., Bustamente, C., & Maestre, M. F. (1980) Ann. Rev. Biophys. Bioeng. 9, 107-142.

Tjerneld, F., Norden, B., & Ljunggren, B. (1979) Photochem. Photobiol. 29, 1115-1118.

Urbanke, C., Romer, R., & Maass, G. (1973) Eur. J. Biochem. 33, 511-516.

Wiesehahn, G. & Hearst, J. E. (1978) Proc. Nat. Acad. Sci. USA 75, 2703-2707.

Young, P. R. & Kallenbach, N. R. (1978) J. Mol. Biol. 126, 467-479.

Youvan, D. C. & Hearst, J. E. (1982) Anal. Biochem. 119, 86-89.

Chapter 4:

Crosslinking of RNA by Different Psoralen Derivatives

INTRODUCTION

The aspects of RNA primary and secondary structure which affect psoralen reaction have been dealt with in the preceding chapters. Equally important in optimizing crosslinking conditions is the effect of psoralen structure. As mentioned in Chapter 1, there are a number of aspects of the psoralen reaction which determine the total amount of crosslink. Previous studies have examined these features (water solubility, photobreakdown rate, photoaddition rate, binding constants) in detail (Isaacs et al., 1977; 1982).

Different substituents on the psoralen ring have dramatic effects on the dark and light reactions. For example, 8-methoxy derivatives have a much lower photobreakdown rate than other psoralens (Isaacs et al., 1982) while trimethyl derivatives yield a much lower percentage of coumarin adducts (Kanne et al., 1982). Understanding of these, and other phenomena, will require analysis of a much larger number of derivatives.

The purpose of this particular work is not to gain a detailed knowledge of how the various derivatives react, but simply to see which are most efficient at generating crosslinks and whether different derivatives produce a different set of crosslinks. This was accomplished by electrophoresing crosslinked RNA in denaturing conditions. The changes

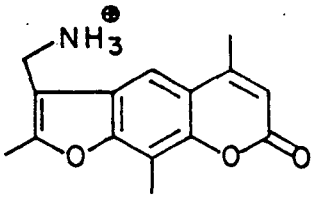
in electrophoretic mobility induced by crosslinking of RNA by psoralen are diagnostic for each crosslink and thus various psoralen derivatives (or other crosslinkers) can be quickly assayed for both efficiency and type of crosslink.

MATERIALS AND METHODS

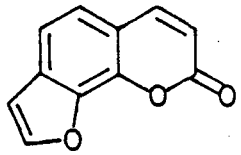
Drosophila melanogaster 5S RNA was prepared and labelled as in Chapter 5. Psoralen derivatives used were synthesized as in Isaacs et al. (1977, 1982). Gels containing 12% polyacrylamide were made as in Chapter 5. Irradiations at 360 nm were done as in Chapter 2. Length of irradiation varied for the different derivatives and was long enough so that greater than 90% of the original psoralen had either covalently added to the RNA or had broken down. Irradiations at 488 nm were done with a Spectra Physics Model 164 Argon Ion Laser using the 488 nm line at 1.8 W for 20 min. During this time, all of the compound (AZT) had broken down or added to the RNA. Samples were irradiated in a water jacketted cuvette at 5° to minimize heating effects.

RESULTS

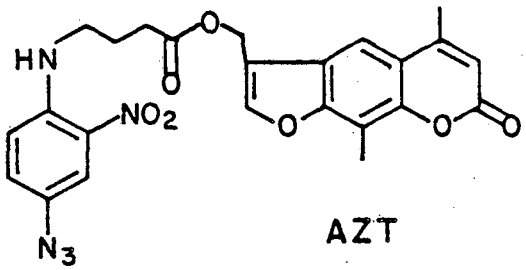
The psoralen derivatives used in this study are pictured in Figure 1. The proper chemical names are listed in the table of abbreviations. The derivatives, which have a wide range of substituents, can be put into two general categories. The first group consists of those which can only covalently react in a manner identical with the parent psor-



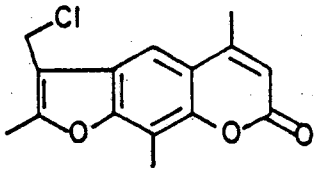
AMT



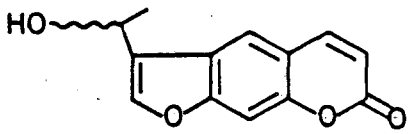
ANG



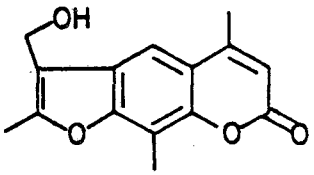
AZT



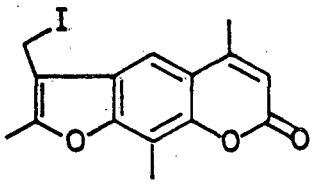
CMT



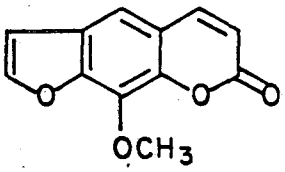
HEP



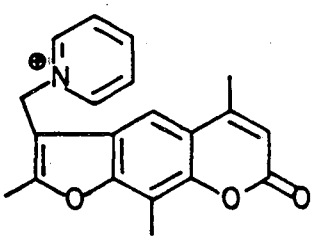
HMT



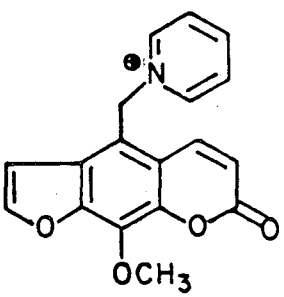
IMT



8 MOP



PMT



PMX

Figure 1: Structure of psoralen derivatives. The chemical structures of all the psoralen derivatives referred to in this chapter are shown. Proper chemical names are listed in the table of abbreviations.

alen, i. e. through the formation of a cyclobutane bridge involving the 4', 5' or the 3, 4 double bond. The second group, made up of CMT, IMT, and AZT, also contain other functionalities which can react with other parts of the RNA. IMT and CMT are alkylating reagents and will react in the dark with the N7 of guanine or the N3 of adenine. The hydrolysis of IMT or CMT to HMT competes with the alkylation. At 0°, all of the psoralen has either reacted with the RNA or hydrolyzed within one minute. CMT is about five times more reactive with poly(G) than with poly(A) while poly(C) and poly(U) are virtually non-reactive. tRNA and 5S RNA are only slightly reactive, despite the presence of potentially reactive purines (see also Chapter 7). AZT contains a phenylazide group which, upon excitation at 450-500 nm, decomposes to a highly reactive nitrene which can insert into any C-H bond. The electronic transitions of the psoralen and phenylazide chromophores do not overlap so either one can be selectively excited without destroying the other.

In lanes 3 and 12, the reaction of HMT with Drosophila 5S RNA is shown (Figure 2). Two major and one minor band can be seen in addition to the normal 5S band. These additional bands represent molecules containing different crosslinks. It is possible that the minor band actually contains two crosslinks (those found in the major bands). None of the psoralen derivatives of the first group yield crosslinks other than those found in these three bands. 8-MOP and AZT (when reacted only at 360 nm so the azido group is not ac-

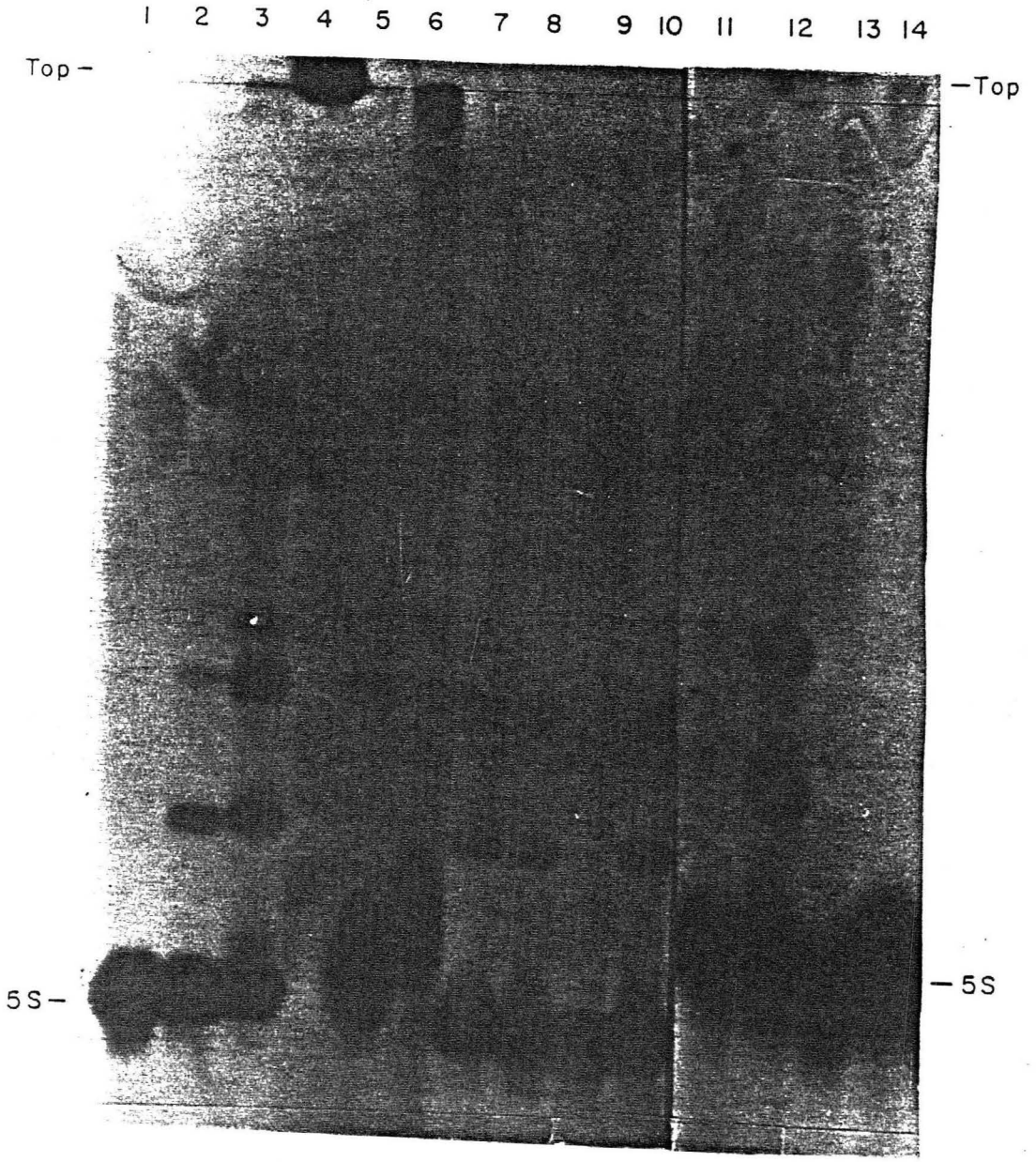

1 2 3 4 5 6 7 8 9 10 11 12 13 14
Top
Top
5S
5S

Figure 2: Mobility of crosslinked 5S RNA. Drosophila melanogaster 5S RNA was crosslinked with the compounds shown and then run on a 12% polyacrylamide, 7 M urea gel at high temperature. The xylene cyanol marker dye migrated through the gel twice. The compounds used, the length of irradiation, and concentrations are listed below.

Lane	Compound	Concentration (μg/ml)	Time of Irradiation (min) 360 nm	488 nm
1	ANG	40	20	-
2	8-MOP	40	120	-
3	HMT	20	30	-
4	AMT	100	30	-
5	PMX	100	120	-
6	PMT	100	120	-
7	AZT	40	20	-
8	AZT	40	20	20
9	AZT	40	-	20
10	AZT	40	20	20
11	HEP	100	120	-
12	HMT	30	20	-
13	IMT	30	20	-
14	AMT	30	20	-

tivated, lane 7) give the major crosslinks, albeit at reduced yield. ANG and HEP samples do not show any change from control samples. The positively charged derivatives show two types of behavior. PMX yields a small amount of one of the major crosslinks while the remainder of the 5S has a small but noticeable retardation in mobility. This is caused simply by the neutralization of the phosphate negative charges by the PMX bound. PMT, on the other hand, yields none of the major crosslinks and has a more pronounced effect on the mobility of the 5S band. Clearly, there are at least a few PMT molecules bound to each 5S as monoadducts. There is also a small amount of material which does not enter the gel. This same phenomena is observed with AMT at low doses (lane 14). At higher doses (lane 4), none of the material enters the gel. Because of the size of 5S RNA, it seems unlikely that a simple crosslink would prevent entrance into the gel. More likely, two or more 5S RNA molecules are crosslinked together. The positive charge of the AMT (or PMT) neutralizes the normally repulsive interaction between phosphates and facilitates close approach. Indeed, Rabin and Crothers (unpublished results) have found that, using AMT, the stems of two 5S molecules can stack on each other and be covalently linked.

When AZT is irradiated only at 488 nm, no crosslinks are observed. If all the unbound AZT is then removed from this sample and irradiated at 360 nm, the two major crosslinks are observed. If the sample which was irradiated

first at 360 nm is subsequently reacted at 488 nm, no new crosslinks are observed, but the pre-existing ones increase slightly in intensity. If IMT is reacted thermally, unbound drug removed, and irradiated at 360 nm, none of the major crosslinks are observed, but a small amount of a new crosslink can be seen.

DISCUSSION

Electrophoresis of Crosslinked Nucleic Acids: Anomolous mobility of psoralen crosslinked DNA (Cech, 1981) and RNA (Wollenzein and Cantor, 1982) in gel electrophoresis has been observed previously. Cech found that the mobility of crosslinked DNA was dependent on the number of crosslinks, the size of the DNA, and the per cent agarose in the alkaline gels. In heavily crosslinked samples, the apparent molecular weight of all sizes of DNA was approximately 1.35-1.53 times that of uncrosslinked DNA and was independent of agarose concentration. Because the psoralen holds together complementary strands even in the absence of base-pairing, the crosslinked DNA should have twice the molecular weight of single stranded DNA. In heavily crosslinked samples, the DNA is constrained to a much smaller volume than normal. The size of the loops caused by unpaired bases is very small compared to the total length of DNA thus it can migrate through the pores in the gel more easily than a single-stranded molecule of equivalent molecular weight. In samples which have only a single crosslink, the behavior is more

complex. DNA up to about 2 kb has an apparent molecular weight of 2.0±0.1 times the single stranded molecular weight. This result shows that a single crosslink has no appreciable effect on the volume of DNA. As the size of the DNA increases above 4 kb, the apparent molecular weight increases dramatically. This occurs because the DNA size has approached the pore size of the gel. Many of the pores present in the gel are simply too small for the DNA to pass through. A similar phenomenon was observed by Mickle et al. (1977) for relaxed circular DNA. At approximately 18 kb, the circular DNA could not enter agarose gels. This contrasts with the results of Roberts and Friedlos (1982) who found that large DNA crosslinked by cisplatin migrated in an alkaline sucrose gradient just as would be expected for its molecular weight. When there is either a large pore relative to the size of the DNA (Cech, 1981) or there is no pore, even for very large DNAs (Roberts and Friedlos, 1982), no alteration in mobility occurs.

At first glance, it would appear that a contradiction arises when it is stated that, in some situations, crosslinked DNA has the same volume or apparent molecular weight as normal DNA while, at other times, there is a large difference in the effective volume. In fact, the dynamic nature or randomly coiled DNA readily accounts for this. Whenever uncrosslinked DNA encounters the gel lattice, it can easily change its shape to move with the voltage gradient. Crosslinked DNA, on the other hand, is covalently fixed. X

structures, arising from double stranded DNA, or loop structures, arising from single stranded RNA, cannot as easily deform and thus cannot migrate as readily.

Wollenzein and Cantor (1982) took advantage of these gross changes in mobility to separate different crosslinked structures in 16S RNA. The relation between loop size and mobility is very complex and will require much more analysis before the migration of a given crosslink can be predicted. Intuitively, larger loops should cause larger mobility changes. The lengths of the tails of the loop also play a major role, however. Longer ends would make it more difficult for a molecule to backtrack if it was trapped in a pore too small for it to pass through.

All of the studies mentioned thus far have used conditions so extreme (alkali or 98% formamide) that no base-pairing could possibly exist. When crosslinks are run under milder conditions (e. g. 7 M urea), great care must be taken in interpreting the presence of secondary structure. Even uncrosslinked secondary structure can exist in the standard 7 M urea polyacrylamide gel (Frank et al., 1981). The fact that a crosslink can stabilize a normally unstable helix has been used to advantage in isolating and analyzing crosslinked fragments (Zwieb and Brimacombe, 1980; Turner et al., 1982; Chapter 6). In chapter 6 and Zwieb and Brimacombe (1980), crosslinks were used to stabilize helices in mild conditions and then fragments were isolated by electrophoresing in denaturing conditions. Turner et al. (1982) found

that psoralen crosslinks stabilized hairpin structures in 7 M urea. When the crosslinks were reversed and the material run under the same conditions (7 M urea, 6°), hairpin structures which had been stabilized by crosslinks became denatured and migrated more slowly in the gel.

The effect of psoralen monoadducts on mobility is much simpler. An HMT monoadduct attached to short oligonucleotides retards the mobility by the equivalent of two nucleotides (Chapter 5). Longer pieces of RNA are similarly retarded (Bachellerie and Hearst, 1982). Presumably this occurs because of hydrophobic interactions between the HMT and the gel lattice. Positively charged derivatives would further slow down any RNA fragment.

While the above data show that the positions of crosslinks cannot be predicted based on mobilities in a gel, it is clear that different crosslinks can be resolved. Thus, with proper choice of acrylamide concentration and care to ensure denaturing conditions, crosslinking within any RNA by any reagent, psoralen or otherwise, can be assayed.

Crosslinking Properties of Different Psoralen Derivatives: As shown in Figure 2, it is clear that HMT is the best crosslinking agent for Drosophila 5S RNA. Other RNAs have not been assayed systematically so it is possible that other RNAs might give different results. However, crosslinking of Drosophila 18S and 5S RNAs follows the same qualitative pattern observed here (Chapter 7) so it seems likely that similar results would be obtained with different systems.

While the yield of crosslinks generated by various derivatives approaches that of HMT in some cases, no other derivative which reacts solely through the normally reactive double bonds produces any different crosslinks. Even when AZT is reacted through its azido group and unreacted material removed, the same pattern of crosslinks emerges. Since the azido group should be virtually non-specific in what it reacts with, this suggests that there are a small number of available intercalation sites and it is the nature of these sites that determine where crosslinking can occur. One would expect any simple intercalator to bind to these same sites; so, in order for a different intercalative crosslinker to generate new crosslinks, it would have to have some other base specificity.

IMT is the only other derivative with a different base specificity and, indeed, it produces one new crosslink. Because it is so specific in its reaction, it must intercalate first and then react once bound. It should be of interest to determine where reaction is occurring because single-stranded guanines are most susceptible to alkylation while intercalation is specific for double-stranded regions. It is not obvious from the secondary structure shown in Chapter 5 where such a crosslink might occur.

Clearly, some substituents cause a profound decrease in crosslinking. Destroying the linearity of the heterocycle, as in ANG, completely abolishes crosslinking. It is possible that tertiary structure in some molecule might produce a

site that is crosslinkable, but 5S RNA is certainly not one of these. Introduction of a positively charged group at the 4' position increases solubility and the amount of monoaddition but drastically reduces crosslinking. Monoaddition of AMT and PMT occurs at a high rate as evidenced by the retardation of the main band of 5S. However, the interaction between the positively charged nitrogen and the phosphate group must be sufficient to pull the psoralen away from the opposite side of the helix and prevent crosslinking. The diameter of the A form helix is so large that any skewing of the psoralen prevents simultaneous alignment of the 4', 5' and 3, 4 double bonds. The B form helix is not as wide so crosslinking of DNA should not be as severely affected. Having the positive charge at the 5 position (as in PMX) lowers the amount of monoaddition but allows crosslinking to occur more readily. Both of these suggest that the interaction between the nitrogen and phosphate is not as strong in PMX as in AMT and PMT.

While no other studies have examined the ability of various psoralen derivatives to form particular crosslinks, there is precedence for the variation of substituents causing changes in the types of products formed. Trimethylpsoralen derivatives have very large photobreakdown rates when compared to 8-methoxy compounds (Isaacs et al., 1982). Even more striking is the effect of the 4-methyl group on the types of monoadducts created. Derivatives containing this group form very small amounts of coumarin adduct (2-3%) in

DNA while derivatives lacking it have up to 20% coumarin adduct. Once again, this seems to be dictated by the structure of the intercalation complex. Similar changes in the amount of coumarin adduct are observed using HMT in different RNA conformations. In DNA, the 4-methyl group must force the psoralen far enough away from the 5, 6 bond of the pyrimidine to inhibit reaction.

The assay described in this chapter provides a way to reliably test compounds for crosslinking ability; and, even more importantly, to determine if crosslinks are produced in different locations. Before investing a good deal of time in analyzing the secondary structure of an RNA by crosslinking, it is certainly advantageous to know if any crosslinks are, in fact, being generated and whether they are the same as any which have been previously identified.

References

Bachellerie, J. -P. & Hearst, J. E. (1982) Biochemistry 21, 1357-1363.

Cech, T. (1981) Biochemistry 20, 1431-1437.

Frank, R., Müller, D., & Wolff, C. (1981) Nucl. Acids Res. 9, 4967-4979.

Isaacs, S. T., Shen, C. -K. J., Hearst, J. E., & Rapoport, H. (1977) Biochemistry 16, 1058-1064.

Isaacs, S. T., Chun, C., Hyde, J. E., Hearst, J. E., & Rapoport, H. (1982) in Trends in Photobiology (Hélène et al., eds.), pp. 279-294, Plenum Publishing Corp.

Kanne, D., Straub, K., Hearst, J. E., & Rapoport, H. (1982) J. Am. Chem. Soc., in press.

Mickle, S., Arena, V. Jr., Bauer, W. (1977) Nucl. Acids Res. 4, 1465-1482.

Roberts, J. J. & Friedlos, F. (1982) Chem. Biol. Interact. 39, 181-189.

Turner, S., Thompson, J. F., Hearst, J. E., & Noller, H. F. (1982) Nucl. Acids Res. 10, 2839-2849.

Wollenzein, P. L. & Cantor, C. R. (1982) J. Mol. Biol. , in press.

Zwieb, C. & Brimacombe, R. (1980) Nucl. Acids Res. 8, 2397-2411.

Chapter Five:

Secondary Structure of *Drosophila melanogaster* 5S RNA by Psoralen Crosslinking

INTRODUCTION

Since the discovery of 5S RNA as a ribosomal component (Rosset and Monier, 1963), considerable work has been done on this molecule (reviewed by Monier, 1974; Erdmann, 1976). Although more than 50 prokaryotic and eukaryotic 5S RNAs have been sequenced (Erdmann, 1980), the secondary structure of the 120 nucleotide long RNA is still the subject of controversy. The specific function of 5S RNA also remains open to question. Activity of reconstituted 50S ribosomal subunits in *E. coli* is dependent upon the presence of 5S RNA (Dohme and Nierhaus, 1976), but deletions and changes in the sequence still allow protein synthesis to occur, albeit at a reduced rate (Pace et al., 1982). In addition, there is suggestive evidence that 5S RNA may be involved in different aspects of ribosomal functions, but none of the evidence at present seems to be conclusive (Erdmann, 1976).

Unlike transfer RNA, for which the correct secondary structure was inferred from the first primary structure (Holley et al., 1965), no definite secondary structure has yet been assigned to 5S RNA. The only common feature among the models proposed to date is the presence of a stem formed by the pairing of the 3' and the 5' ends of the molecule. A model which relies on the comparison of all avail-

able sequences and which considers only perfect pairings has been proposed by Fox and Woese (1975). Vigne and Jordan (1977) have proposed a very similar model based on partial enzymatic digestion experiments performed on 5S RNA from a variety of organisms. They have shown that two regions of 5S RNA (around nucleotides 40 and 90) are particularly accessible to ribonucleases and, thus, are most probably single-stranded. Laser raman spectroscopy data, which suggest that some 60% of the bases are paired in 5S RNA, have led Luoma & Marshall (1978a,b) to propose another model very similar to the "cloverleaf" structure of tRNA. Other techniques have also been employed to study single-stranded regions of the molecule and to determine the number and type of base-pairs (Monier, 1974; Erdmann, 1976). These data are of limited usefulness, however. More detailed information about specific intramolecular interactions is needed.

In the present study, we have attempted to get direct information about the secondary structure of 5S RNA by crosslinking with HMT. Three other crosslinking studies have already been done on 5S RNA from Escherichia coli. One study was with a derivative of psoralen, aminomethyl-trioxsalen (Rabin & Crothers, 1979), while the other two were with 1,4-phenyl-diglyoxal (Wagner & Garrett, 1978; Hancock & Wagner, 1982). The latter reagent has an unknown specificity because glyoxal reacts preferentially with single-stranded residues while the phenyl ring might

allow intercalation. Because of its specificity, psoralen is highly preferable in this regard. The crosslinks identified agree very well with the model of Fox & Woese (1975) and Vigne & Jordan (1977). In addition, we propose modifications which now account for the raman data of Luoma & Marshall (1978a). This work was done, in part, with Maurice Wegnez, and appeared in Journal of Molecular Biology, volume 147, pp. 417-436.

MATERIALS AND METHODS

(a) Purification and labelling of 5S RNA

i) 5S RNA from Drosophila embryos

Dechorionated Drosophila embryos (20 to 25 g) were homogenized in 250 ml of 200 mM sodium acetate (pH 5.0). RNA was extracted with phenol in the presence of 0.5% (w/v) sodium dodecyl sulfate at 0°C (Brown & Littna, 1964). After precipitation with ethanol, RNA was dissolved in 1.2 M NaCl, 50 mM ammonium acetate (pH 5.3), 5 mM $MgCl_2$ and kept overnight at 0°C. Cold precipitated high molecular weight RNA was removed by centrifugation and the supernatant was diluted to 100 mM NaCl. Whatman DE52 cellulose was added and washed several times with 100 mM NaCl, 50 mM ammonium acetate (pH 5.3), 5 mM $MgCl_2$. RNA was then eluted by raising the salt concentration to 1.2 M NaCl with a yield of 800 O.D. units at 260 nm. 5S RNA, accounting for 20% of this preparation, was purified through two cycles of Sephadex G100 chromatography (Wegnez et al., 1978).

(ii) 5'-End labeling of 5S RNA

5S RNA (500 µg) was first dephosphorylated in 200 µl of 50 mM Tris-HCl (pH 9.0) with 1.2 units of calf alkaline phosphatase (Boehringer) at 37° for 30 min. The RNA was extracted twice with phenol, then purified on a 12.4% (w/v) polyacrylamide/8 M urea gel polymerized in 40 mM Tris-acetate buffer (pH 8.3). The full length 5S RNA band was excised, eluted with 0.3 M NaCl and precipitated with ethanol. Phosphorylation of 5S RNA (2 µg) was carried out in 50 µl of 50 mM Tris-HCl (pH 9.5), 10 mM $MgCl_2$, 5 mM dithiothreitol containing 150 µCi [γ-^{32}P]ATP (Amersham; 3000 Ci/mmol) and 1 unit of T_4 polynucleotide kinase (Boehringer). The incubation was carried out at 37° for 30 min and was terminated by extraction with phenol. Purification of full length ^{32}P-5S RNA on a 12.4% polyacrylamide/8 M urea gel provided approximately 10^7 cts/ min of activity.

(iii) 5S RNA labeling from cell culture

5S RNA uniformly labeled with ^{32}P was prepared by a modification of the method of Jordan et al. (1976). Actively growing KC Drosophila cells (Echalier & Ohanessian, 1970) were transferred from complete medium to low phosphate medium (obtained from the cell culture facility, University of California, San Francisco) and allowed to grow 24 h at 25°C. ^{32}P in the form of orthophosphate (Amersham; 8 mCi/ mmol) was then added to a concentration of 0.1 mCi/ml and the cells were harvested after 48-72 h of growth. RNA was extracted by the sodium dodecyl sulfate/cold phenol

method of Brown & Littna (1964) and loaded on a non-denaturing 12.4% polyacrylamide gel in 40 mM-Tris-acetate buffer (pH 8.3) (Benhamou et al., 1977). The 5S RNA band was eluted with a small volume of 0.3 M NaCl and precipitated with ethanol. Typically, 100 ml cultures yield 1.5 x 10^6 to 4.0 x 10^6 cts/min of 5S RNA.

(b) Partial digestion of 5S RNA

Partial digestions of 5S RNA with T_1 ribonuclease were performed at 0 to 4°C in different salt concentrations. After the digestion (3 to 5 min), 5S RNA was extracted with phenol, precipitated with ethanol, dissolved in 20 to 40 µl of 40 mM-Tris-acetate (pH 8.3), 8 M urea and analyzed on a 12.4% polyacrylamide gel made up in the same buffer. After electrophoresis, the gel was shaken 10 to 15 minutes in electrophoresis buffer and then stained with ethidium bromide (1 µg/ml) and/or autoradiographed.

(c) Crosslinking

The 5S RNA was dissolved in 0.4 ml of 5 mM NaCl, .2 mM Tris·HCl (pH 7.5), 0.02 mM EDTA. [^{3}H]HMT (3.7 x 10^7 cts/min per µg, a generous gift of Steve Isaacs) was synthesized according to Isaacs et al. (1982) and added at concentrations of up to 30 µg/ml, depending on the level of incorporation desired. The solution was irradiated in 1.5 ml Eppendorf tubes. Two or 3 cycles of 5 min irradiations separated with new additions of HMT were routinely performed at 4° as described in chapter 2.

(d) Sequence techniques

The RNA was digested with 10 µl of T_1 RNase (Sankyo; 5000 units/ml, 50 mM Tris, pH 7.5). The digestion was for 2 h at 37°. After digestion, 5 µl of 7 M urea, 1 M EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol, 50 mM Tris-borate (pH 8.3) was added. The sample was loaded into a 0.8 cm well in a 20% polyacrylamide gel (20% acrylamide, .6% bis acrylamide, 7 M urea, 50 mM Tris-borate, 1 mM EDTA; 0.05 cm x 12 cm x 40 cm) and run at 1000 to 1500 V until the dye markers were separated by 11 cm. After autoradiography, the fragments of interest were cut from the gel, eluted in 1 ml of 0.3 M NaCl, dialyzed against H_2O, lyopholyzed, and resuspended in 10 µl of H_2O. The crosslinks were then reversed with a 6 W hand-held, Rayonet short-wave UV lamp. The lamp was at a distance of 5 cm from the sample. Crosslinks in dilute solution are 50% reversed in 20 min. The reversed fragments were run on a 20% gel and isolated as described above. Base compositions of the isolated fragments were done by further digestion with RNase T_2 (Sigma; 10 µl of 2000 u enzyme/ml in 50 mM sodium acetate, pH 4.5; 16 h at 37°) or RNase A (Sigma; 10 µl of 10 mg enzyme/ml in 50 mM Tris, pH 7.5; 2 h at 37°). The products were then analyzed by paper electrophoresis as described in chapter 2.

RESULTS

The 5S RNA which was crosslinked in this study was pure and homogeneous. Purification on the non-denaturing gel yields a sharp band. Under the same conditions, E. coli 5S

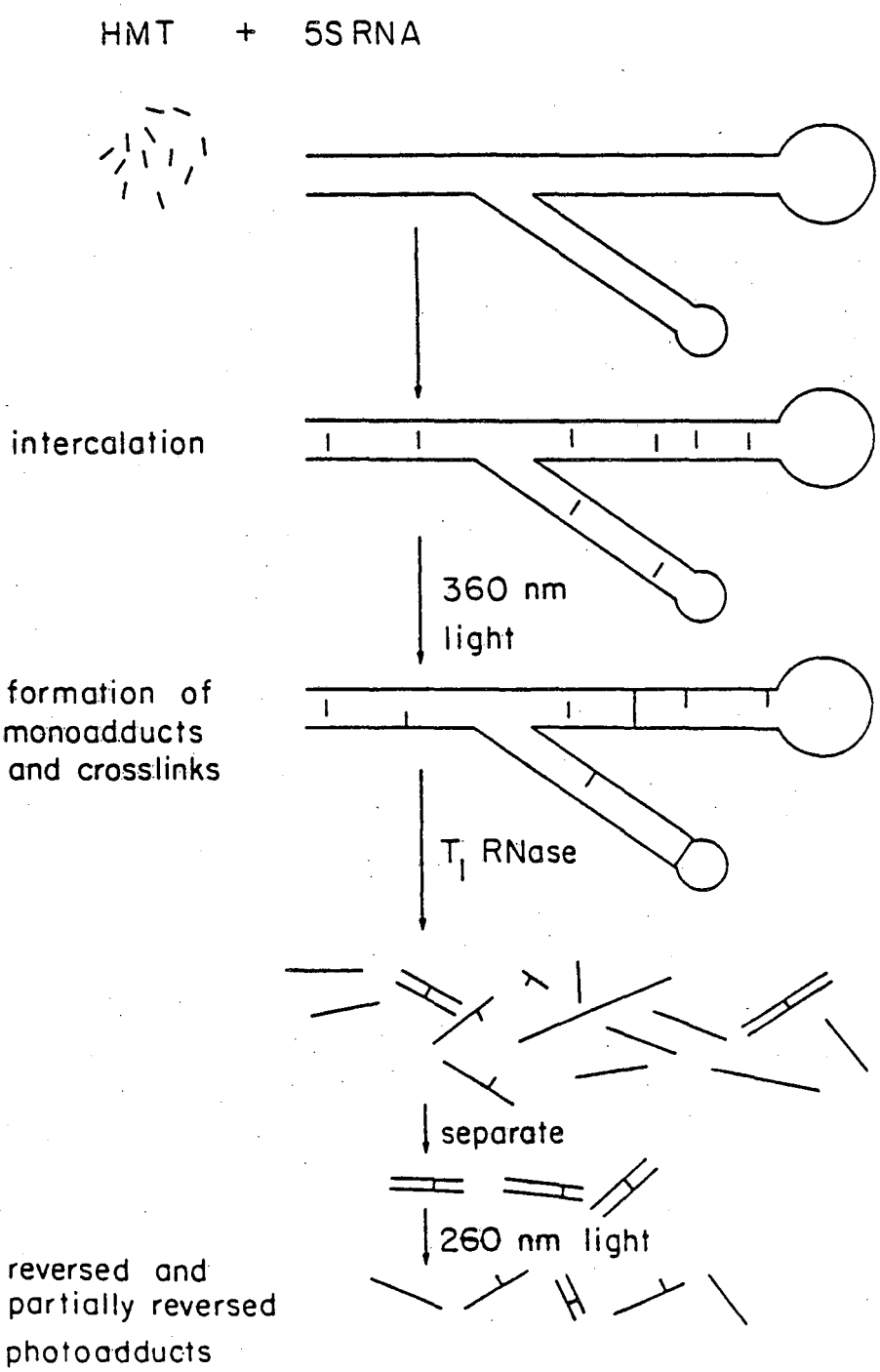

HMT + 5S RNA
intercalation
360 nm
light
formation of
monoadducts
and crosslinks
T1 RNase
separate
260 nm light
reversed and
partially reversed
photoadducts

Figure 1: Schematic diagram for forming and analyzing crosslinks.

RNA runs as two bands (A and B forms described by Aubert et al., 1968; experiment not shown). Luoma et al. (1980) also noted that S. cerevisiae 5S RNA has a homogeneous structure in these conditions. The two dimensional fingerprints obtained after RNase T_1 and RNase A digestion are identical to those published by Benhamou et al. (1977). Samples run under denaturing conditions show that there are no nicks (experiments not shown).

The amount of HMT incorporated into 5S RNA is highly dependent on the temperature and salt concentration used in the irradiation (Fig. 2). Crosslinking was done at low salt concentration and low temperature in order to increase the level of addition. The effects of these irradiation conditions on the conformation of 5S RNA was determined by partial hydrolysis experiments.

Drosophila 5S RNA is cut after residues 37 and 89 when subjected to a mild T_1 RNase digestion (Benhamou et al., 1977). With the conditions we used for partial hydrolysis, only 3 bands are detected (Fig. 3(a)). The slowly moving band corresponds to intact 5S RNA while the other two bands result from a single cleavage. When 5S RNA labeled at the 5' end is used, only the two slowly moving bands can be seen by autoradiography (experiment not shown). These two bands correspond to intact 5S RNA and fragment 1-88. The presence of HMT in the incubation mixture does not change the digestion pattern of Fig. 3(b). Additional bands are seen when ethidium bromide is present (Fig. 3(c)). These are caused by the

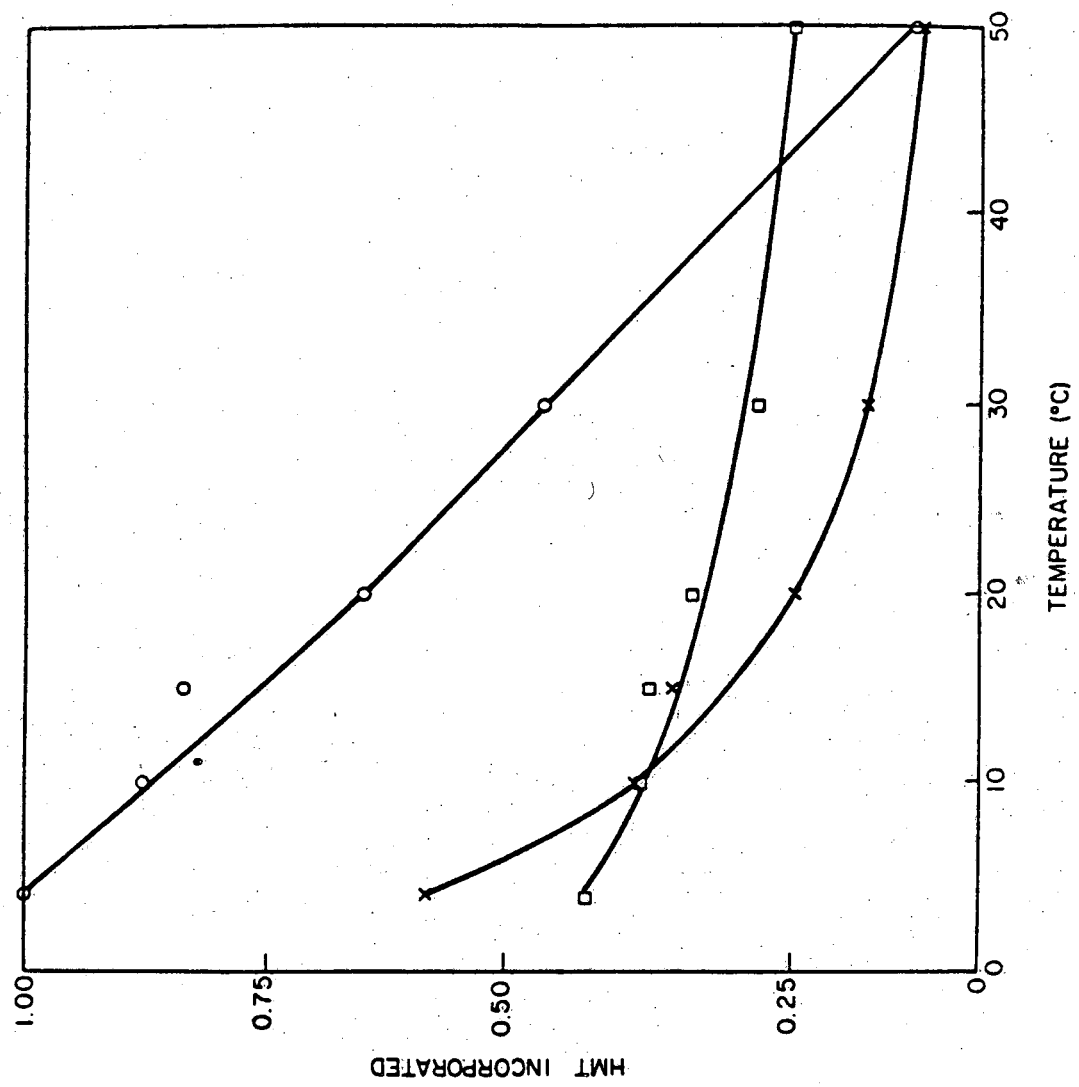

HMT INCORPORATED
1.00
0.75
0.50
0.25
0
0
10
20
30
40
50
TEMPERATURE (°C)

Figure 2: Relative rates of HMT incorporation into 5S RNA as a function of temperature for: 5 mM NaCl (o—o), 100 mM NaCl (□—□), and 5 mM $MgCl_2$, 5 mM NaCl (×—×). All samples contained 20 µg/ml 5S RNA, 1 µg/ml HMT, 2 mM Tris-HCl (pH 7,5), 0.2 mM EDTA and were irradiated for 2 min. Samples were extracted twice with phenol, ethanol precipitated twice and then counted.

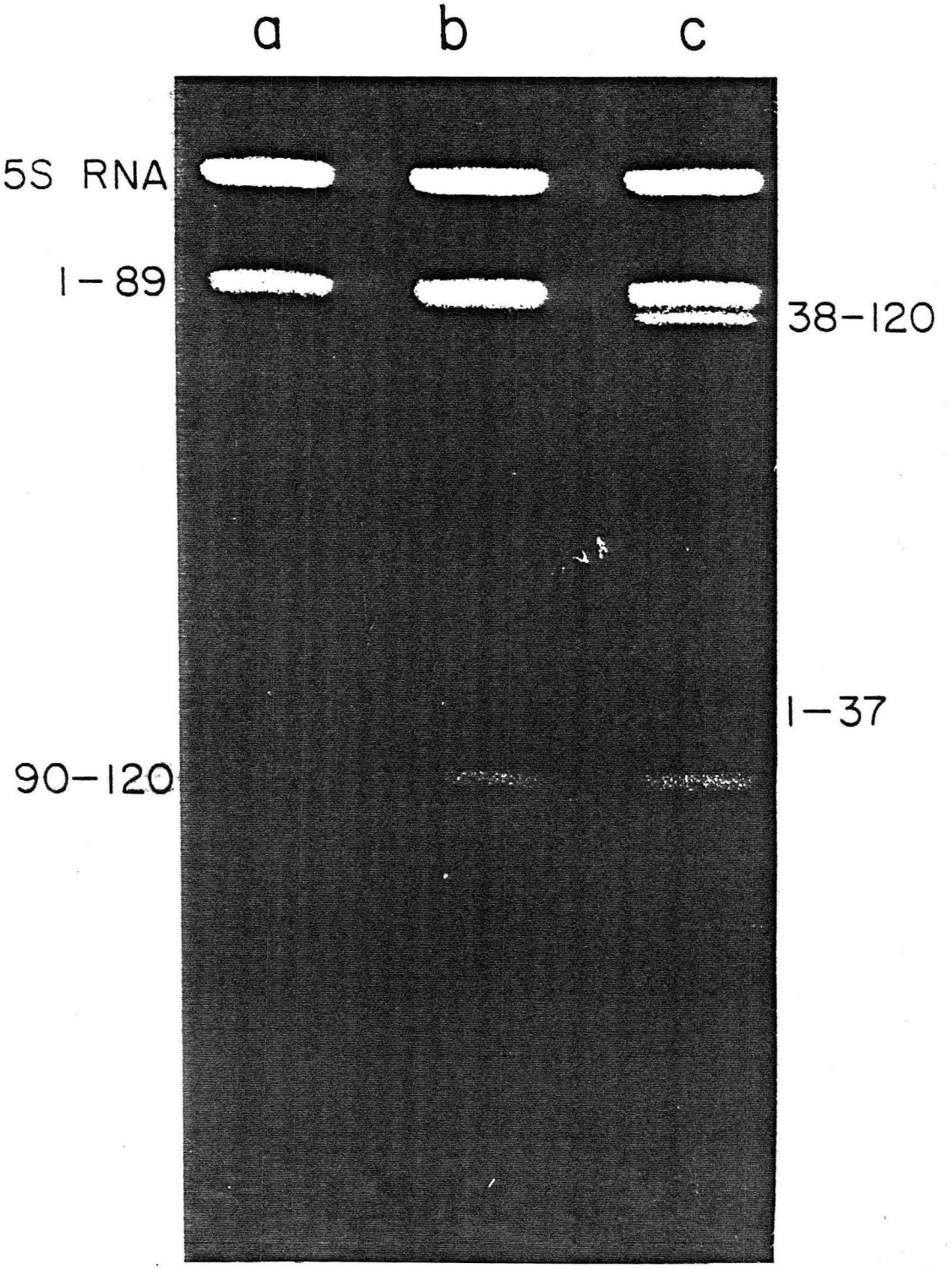

a
b
c
5S RNA
1-89
38-120
1-37
90-120

Figure 3: Partial RNase digestion of Drosophila 5S RNA. 5S RNA (7 µg) was hydrolyzed with 5 units of T_1 RNase in 40 µl of 200 mM NaCl, 20 mM $Mg(C_2H_3O_2)_2$, 50 mM Tris-HCl (pH 7.5), at 0-4° for 5 min (a). In (b), the digestion was done under identical conditions except that HMT (30 µg/ml) was added. In (c), ethidium bromide (30 µg/ml) was added.

the cleavage after G37. The same pattern is also observed with and without Mg^{2+} and over a wide range of Na^{+} concentrations. This indicates that the secondary structure is stable even at very low salt concentrations. The fact that the same crosslinks are produced in low salt and in Mg^{2+} is shown in Figure 4. Migration of RNA in denaturing gels is retarded by the presence of crosslinked loops (see chapters 4 and 6 for further discussion). While the yield of crosslinks varies in the two salt conditions, the pattern of bands is the same.

The types of HMT adducts which are formed upon irradiation can be easily determined by paper electrophoresis. Studies with model compounds (chapters 2 and 3) as well as empirical calculations (Sommer, 1979) have been used to determine the mobilities of all possible monoadducts and crosslinks. A T_2 RNase digestion of 5S RNA reacted with [^{3}H]HMT is shown in Figure 5. One major peak is observed after a 15 second irradiation. This has been assigned to the monoadduct of uridine. Two small peaks can also be seen near the origin. The smaller one is probably a breakdown product of HMT that is not removed in the purification. The other peak is a monoadduct of cytidine. After 15 min, two peaks that migrate faster than UMP are seen. These are isomers of a crosslink between two uridines. Part of the slowest moving peak attributed to the uridine monoadduct may be a uridine-cytidine crosslink. It comigrates with one of the isomers.

Polyacrylamide sequencing gels were used to analyze the

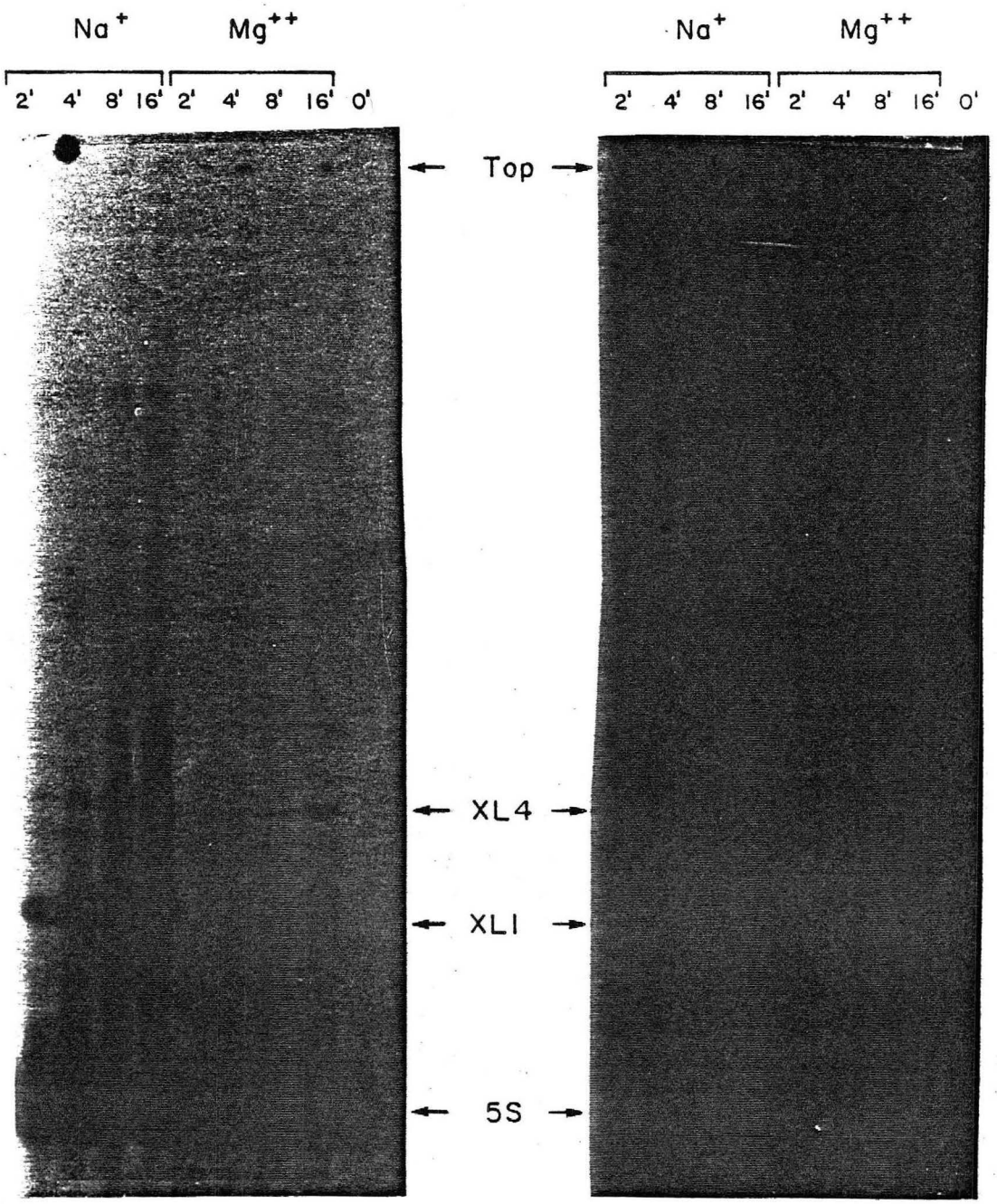

Na+
Mg++
2' 4' 8' 16' 2' 4' 8' 16' 0'
Top
XL4
XL1
5S
Na+
Mg++
2' 4' 8' 16' 2' 4' 8' 16' 0'

Figure 4: Production of crosslinks in Na^+ and Mg^{2+}. Two exposures of the same 20% polyacrylamide gel are shown. Both films are overexposed in order to show minor crosslinks more clearly. 5S RNA which is incubated in the presence of 20 µg/ml HMT but with no irradiation migrates as a single band (0'). Varying irradiation times are shown for Na+ buffer (5 mM NaCl, 1 mM Tris-HCl, pH 7.5, 0.1 mM EDTA) and for Mg^{2+} buffer (200 mM NaCl, 20 mM $Mg(C_2H_3O_2)_2$, 50 mM Tris-HCl, pH 7.5).

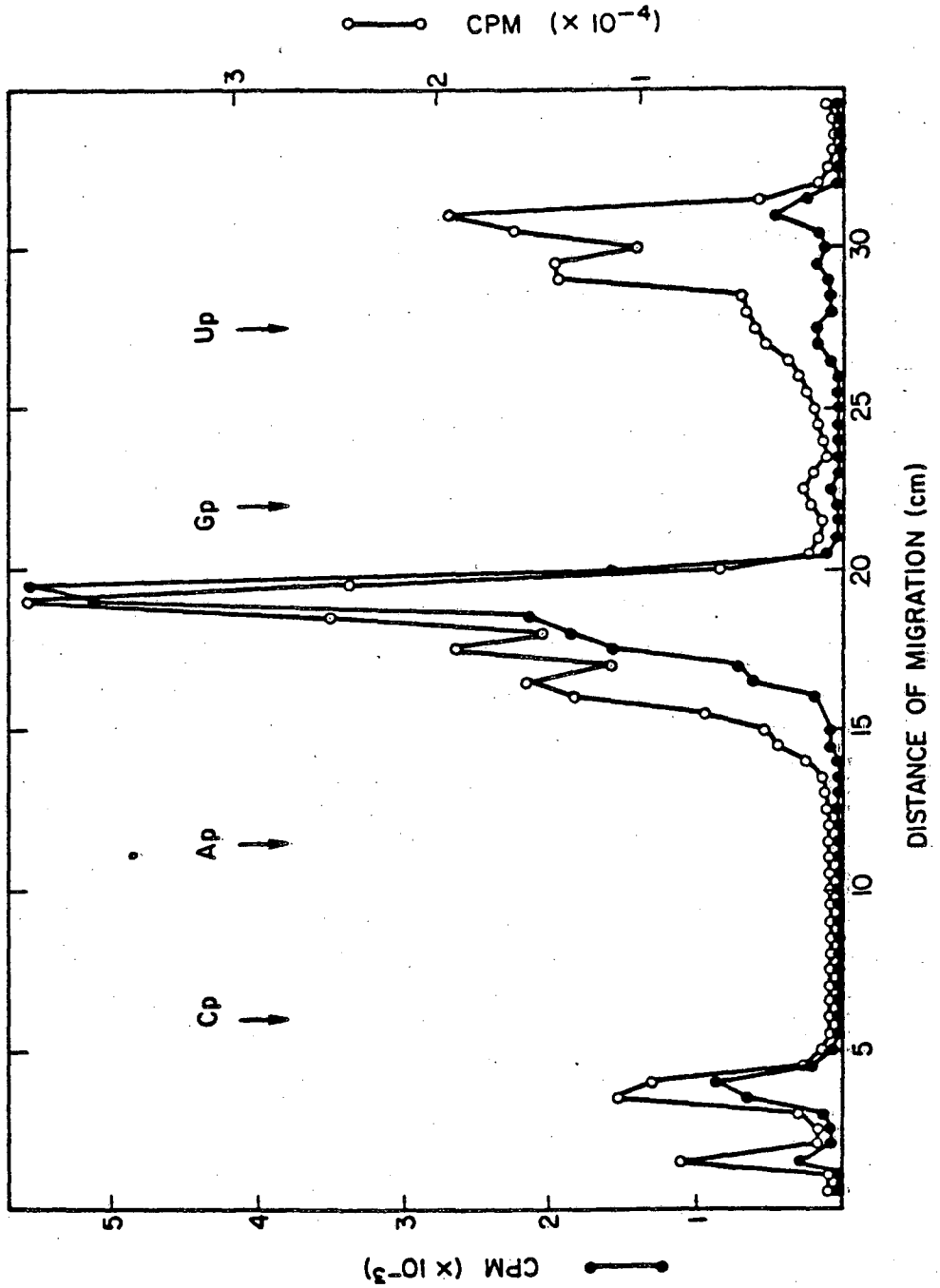

CPM (× 10⁻⁴)
CPM (× 10⁻³)
DISTANCE OF MIGRATION (cm)
Up
Gp
Ap
Cp

Figure 5: Paper electrophoresis analysis of a T_2 RNase hydrolysate of Drosophila ^{32}P-5S RNA which has been reacted with ^{3}H-HMT. The RNA was irradiated for 15 s (●—●) or 15 min (○ - - ○) in the presence of 5 µg/ml HMT. After the paper dried, 0.5 cm slices were cut and counted. Arrows indicate the location of the ^{32}P-mononucleotides while the complete ^{3}H profile is shown.

HMT photoadducts after total T_1 RNase digestion of 5S RNA. As shown in Figure 6, exactly the same pattern of oligonucleotides was obtained when the 5S RNA was untreated (lane A), when HMT was added to the RNA with no light (lane B) and when the RNA was irradiated with no HMT present (lane C). The length distribution of fragments is that expected. The faint, slowly moving bands in lanes A to C probably result from incomplete digestion. Several new bands appear when 5S RNA is irradiated in the presence of HMT. The ^3H profile from the [^3H]HMT in lane D is shown in Figure 7. Most of the bands running slower than the longest unmodified oligonucleotides are expected to be crosslinks.

Attempts to analyze the intact crosslinks met with limited success. The number of possible crosslinks is large and minor contaminants can cause errors in the expected base compositions. To avoid these problems, the crosslinks were reversed prior to analysis. Reversal of cyclobutane-type compounds has been observed previously with psoralen adducts (Musajo et al., 1967; Rabin and Crothers, 1979). After the crosslink between oligonucleotides has been reversed, they can be separated and analyzed independently. The length and base compositions uniquely determine the fragments involved in the crosslinks.

Crosslink 4A (XL4A, Fig. 6) is the easiest to analyze. Upon reversal, bands corresponding to lengths of 8 (XL 4A-A) and 9 (XL4A-B) bases are observed (Fig. 8). There is only one oligonucleotide corresponding to each of these lengths

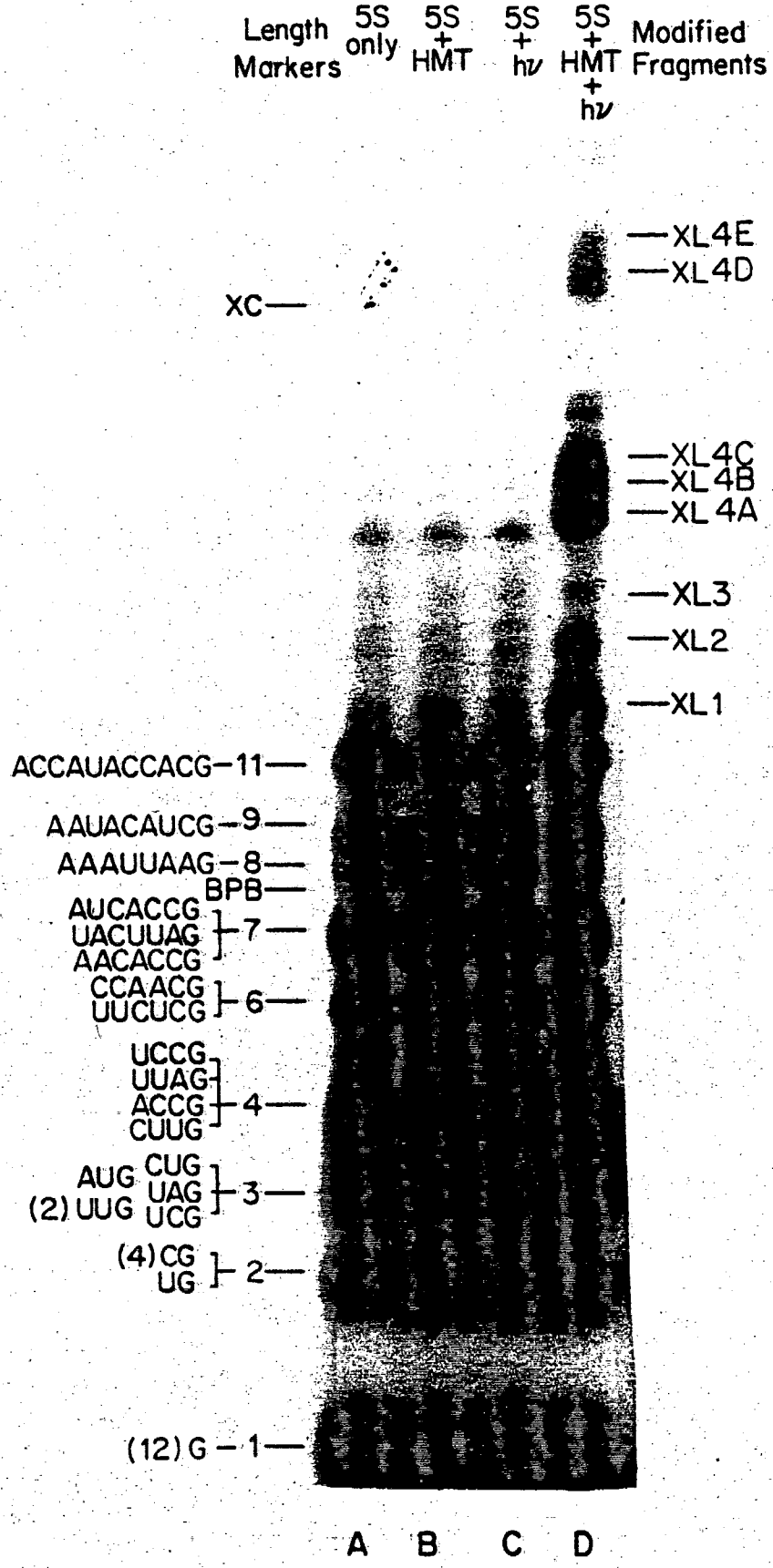

Length Markers
5S only
5S + HMT
5S + hν
5S + HMT + hν
Modified Fragments
XC
XL4E
XL4D
XL4C
XL4B
XL4A
XL3
XL2
XL1
ACCAUACCACG–11
AAUACAUCG–9
AAAUUAAG–8
BPB
AUCACCG
UACUUAG
AACACCG
7
CCAACG
UUCUCG
6
UCCG
UUAG
ACCG
CUUG
4
AUG CUG
(2) UUG UAG
UCG
3
(4) CG
UG
2
(12) G – 1
A
B
C
D

Figure 6: Autoradiogram of T_1 RNase digestion of ^{32}P-5S RNA Samples were run on a 20% polyacrylamide gel after digestion and a) no treatment, b) addition of 20 µg/ml HMT, c) 10 min irradiation, d) addition of 20 µg/ml HMT and 10 min irradiation. The column at left shows the length and sequence of the T_1 fragments as well as the positions of the bromophenol blue (BPB) and xylene cyanol (XC) dye markers. The column at right shows the positions of the crosslinks described in the text.

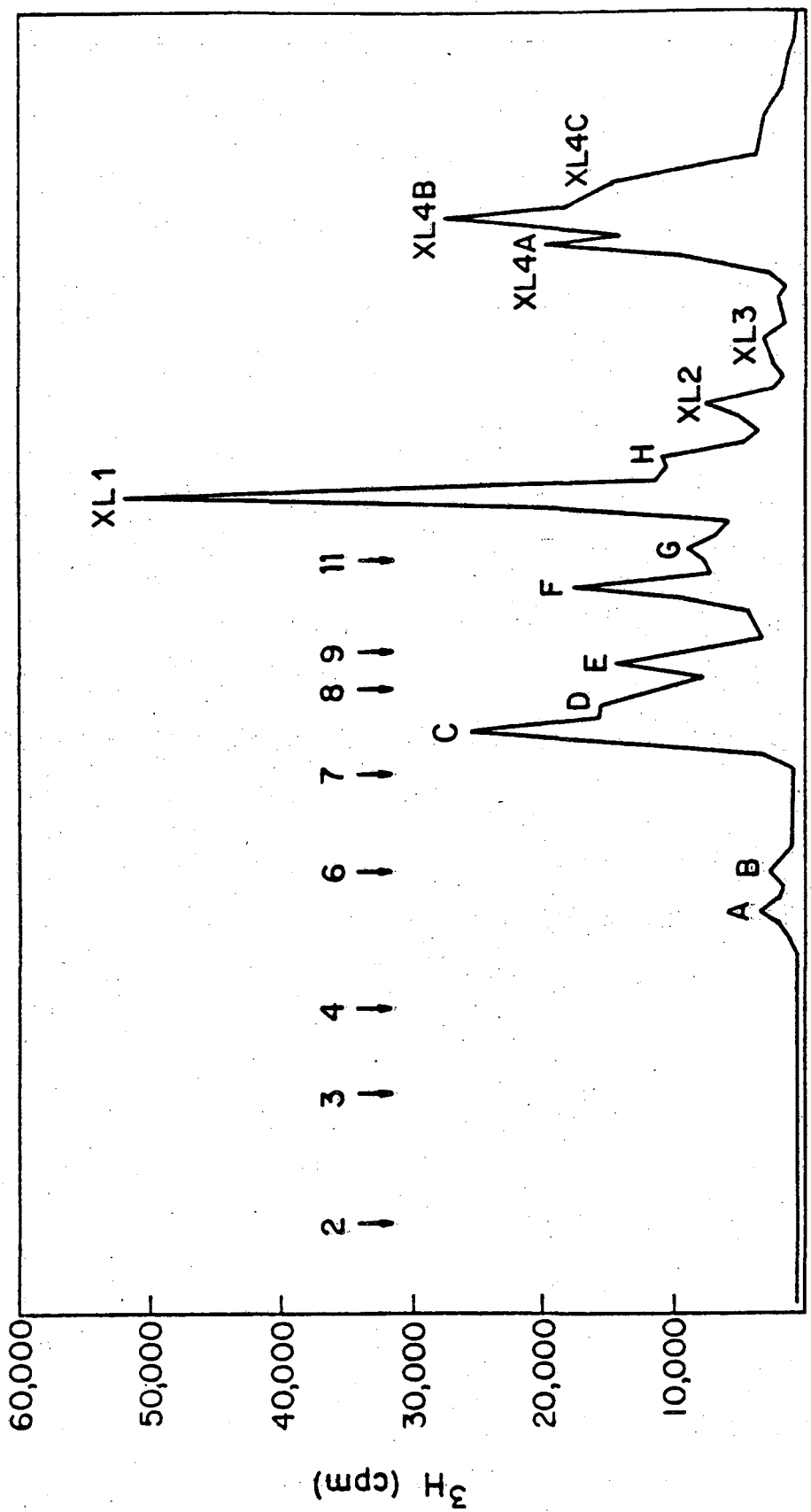

60,000
50,000
40,000
30,000
20,000
10,000
^3H (cpm)
2
3
4
6
7
8 9
11
A
B
C
D
E
F
G
H
XL1
XL2
XL3
XL4A
XL4B
XL4C
PEAK

Figure 7: ^{3}H incorporation into T_1 RNase oligonucleotides. Lane d from Figure 6 was cut into 0.2 cm slices. The RNA was eluted in 0.3 M NaCl overnight and then counted. Arrows indicate the positions of the unmodified oligonucleotides. the crosslinks described in the text are labeled XL1-4 while the other peaks are lettered and identified in Table II.

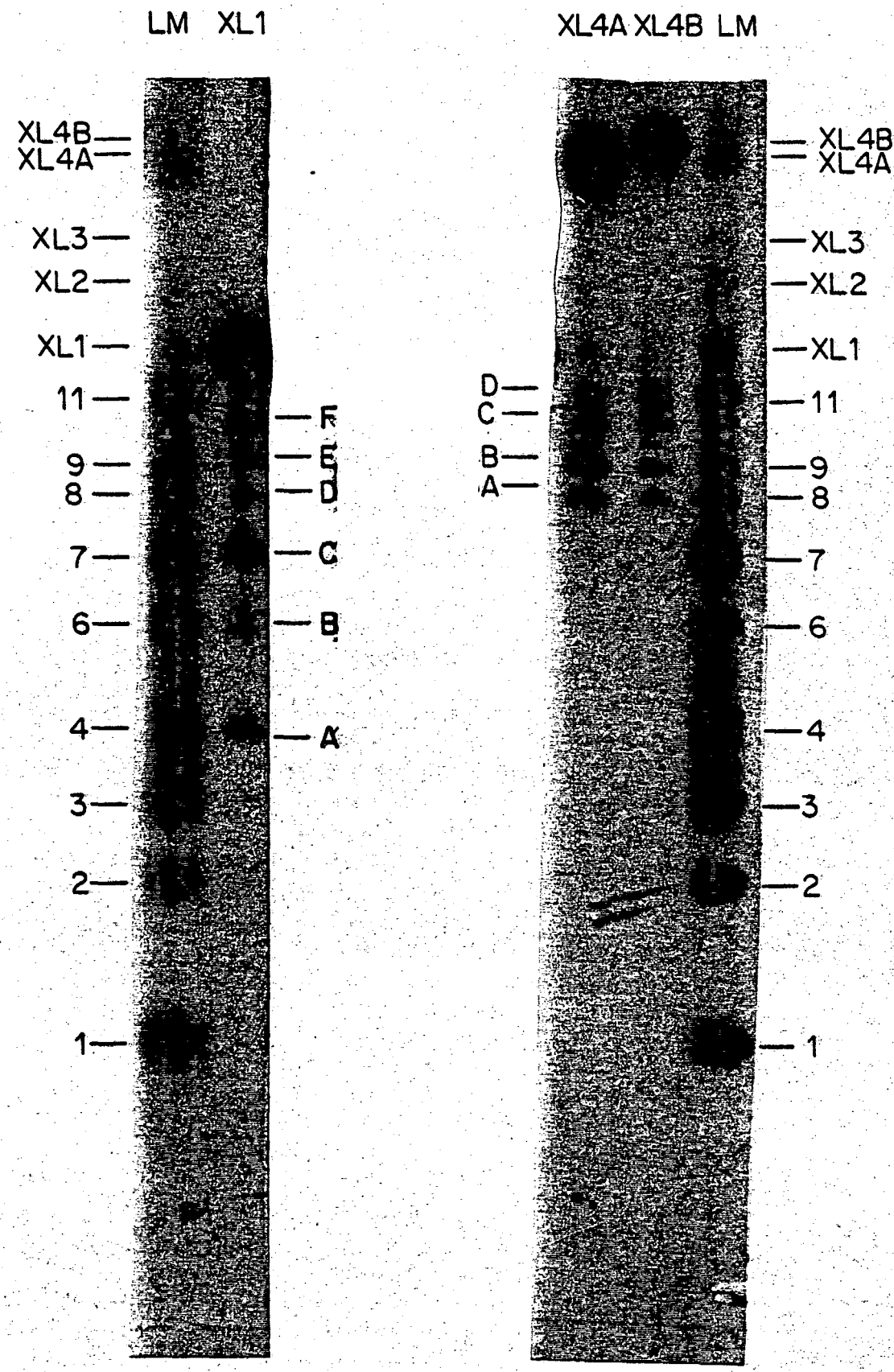

LM XL1
XL4B
XL4A
XL3
XL2
XL1
11
9
8
7
6
4
3
2
1
F
E
D
C
B
A
XL4A XL4B LM
D
C
B
A
XL4B
XL4A
XL3
XL2
XL1
11
9
8
7
6
4
3
2
1

Figure 8: Reversal of crosslinks with short wave UV light. Crosslinks were reversed as described in Materials and Methods. Crosslink 1, crosslink 4A, and crosslink 4B are shown. The lanes at the far left and right contain the complete digest as length markers. The reversal products from each crosslink are lettered and identified in Table I.

Table I: Base Compositions of Crosslinked Oligonucleotides

Fragment	Experimental (theoretical) Base Compositions					Length From Gel	Inferred Sequence	Position in 5S RNA
	C	A	G	U	U*			
XL1-A	0.99(1)	0.00(0)	1.05(1)	1.95(2)	0.00(0)	4	CUUG	94-97
XL1-B	1.08(1)	0.00(0)	0.97(1)	1.10(1)	0.85(1)	6	CU*UG	94-97
XL1-C	1.11(1)	1.80(2)	1.23(1)	2.87(3)	0.00(0)	7	UACUUAG	76-82
XL1-D	1.24(1)	1.95(2)	0.97(1)	1.97(2)	0.86(1)	8	UACUU*AG	76-82
XL1-E	1.65(2)	3.96(4)	1.12(1)	2.26(2)	0.00(0)	9	AAUACAUCG	22-30
XL1-F	0.68(0)	4.33(5)	0.90(1)	1.16(1)	0.92(1)	10	AAAU*UAAG	49-56
XL4A-A	0.00(0)	4.96(5)	1.21(1)	1.84(2)	0.00(0)	8	AAAUUAAG	49-56
XL4A-B	1.85(2)	3.71(4)	1.11(1)	2.33(2)	0.00(0)	9	AAUACAUCG	22-30
XL4A-C	0.34(0)	4.61(5)	1.00(1)	0.97(1)	1.08(1)	10	AAAU*UAAG	49-56

Table II: Assignment of Modified Nucleotides

Peak	Percent of Total ^3H	Length[1] According to $^{32}P/^3H$	Length According to Gel	Assigned Length	Number of HMTs per Fragment	Position in 5S RNA
A	1.48	-	6	4	1	94-97?
B	1.27	-	6	-	-	?
C	13.51	6.39	8	6	1	32-37?
D	5.83	-	8	-	-	?
E	6.86	-	9	7	1	76-82
F	8.39	-	10	8	1	49-56
G	5.06	-	11	9	1	22-30
H	5.64	6.63	13	8	2	49-56
XL 1	17.65	11.00	12	4+7	1	76-82 94-97
XL 2	3.24	13.59	15	6+7	1	32-37 42-48?
XL 3	1.84	14.85	16	3+11	1	8-18 111-113?
XL 4A	8.27	16.12	19	8+9	1	22-30 49-56
XL 4B	11.43	9.84	20	8+9	2	22-30 49-56
XL 4C	9.55	8.40	21	8+9	2	22-30 49-56

[1]The length according to $^{32}P/^3H$ was calculated by setting the length of XL 1 to 11 and comparing all to it. In all cases, fragments were assumed to have one HMT.

in the molecule. These oligonucleotides occur between 22-30 and 49-56. The base composition is exactly that expected (Table 1). There are also two slower moving bands running as 10 (XL4A-C) and 11 (XL4A-D). The base composition of XL4A-C corresponds to the octamer seen above, but with a monoadduct still remaining. XL4A-D could not be obtained in sufficient quantity to analyze. Presumably it corresponds to the monoadduct of the nonamer. In preparations with a high incorporation of HMT, one or two bands (XL4B and XL4C; Fig. 6) are observed running slightly slower than crosslink 4A. Reversal yields the same fragments seen in the main band (Fig. 8). Based on the ratio of ^{3}H to ^{32}P, these have been assigned to the same crosslink with an additional monoadduct (Table 2). Some fainter bands running much slower on the gel have also been attributed to this crosslink (XL4D and XL4E; Fig. 6). They probably result from an incomplete T_1 digestion caused by the bound HMT.

Reversal of crosslink 1 (XL1; Fig. 6) yields fragments corresponding to lengths 4 (XL1-A) and 7 (XL1-C). The base compositions (Table 1) show that these are the oligonucleotides occurring between 94-97 and 76-82. Once again, bands corresponding to oligonucleotides still containing a monoadduct are observed (XL1-B and XL1-D). Two additional bands can also be seen (Fig. 8). The base compositions show these to be the octamer (XL1-E) and nonamer (XL1-F) seen in crosslink 4. These fragments appear because the

octamer with two monoadducts and the nonamer with one monoadduct run very close to crosslink 1. After irradiation with 260 nm light, the monoadducts are lost and the bands migrate faster. Crosslink 2 (XL2; Fig. 6) was more difficult to analyze because of the low yield. Upon reversal (not shown), fragments corresponding to lengths of 6 and 7 bases are observed. There were not enough counts to analyze the base compositions.

An additional crosslink (XL3; Fig. 6) is also present. Upon reversal, a band corresponding to a length of 11 is seen. The length of the total crosslink should be 14 or 15 bases, based on its position. No band of length 3 or 4 is seen, however. Since it would only have 30% of the radioactivity of the long fragment, this is not surprising. Because there is only one fragment 11 bases long in the molecule, the crosslink probably occurs between fragment 8-18 and a tri- or tetranucleotide.

DISCUSSION

(a) Reaction of HMT with 5S RNA

The ability of HMT to intercalate and react with 5S RNA is highly dependent on temperature and the presence of Mg^{2+} (Fig. 2). As also observed with DNA (Hyde & Hearst, 1978) and other RNAs (Chapter 3), Mg^{2+} sharply reduces the uptake of HMT by 5S RNA. This may be explained by the strong stabilization of the double-stranded regions which prevents the HMT from intercalating. The fact that

the incorporation of HMT into 5S RNA is not as dependent on Mg^{2+} as it is in tRNA may indicate less tertiary structure. This has also been suggested on the basis of melting and NMR studies (Luoma et al., 1980). The lowered incorporation of HMT at increased temperatures is caused by the melting of the RNA.

The identities of the fragments in bands A to H (Fig. 7) have not been determined. Most of these are monoadducts, but it has not been possible to purify them for more detailed study. They are not present in as large quantities as the crosslinks.

In any attempt to determine the secondary structure of RNA molecules, a critical examination of the experimental conditions should be done to ascertain their influence on the RNA conformation. As seen in figure 4, the same crosslinks are formed with Mg^{2+} and at low salt. In order to see the effects on the secondary structure of the optimal conditions for HMT uptake by 5S RNA (4°C, low salt concentration, no Mg^{2+}), we made a partial digestion of 5S RNA in those conditions. Partial enzymatic digestions are assumed to attack exposed single-stranded regions preferentially. This is probably true for first cleavages, but subsequent cleavages may be the result of an altered structure caused by first cuts. Figure 3 shows that only one cut, after G88, occurs in 5S RNA following T_1 partial hydrolysis when the high ionic buffer (200 mM-NaCl, 20 mM $Mg(C_2H_3O_2)_2$) described by Vigne & Jordan (1977) is used.

Benhamou et al. (1977) observed two cleavages. We observe the second cleavage (after G37) at a much lower level. While the cleavage at G37 cannot be seen in Figure 3(a) or (b), it can be seen when T_1 is used at slightly higher concentrations. This difference probably is caused by the way in which the sample is handled after the digestion. Benhamou et al. (1977) ran the sample directly on a gel while we first phenol extract and precipitate. This stops digestion more completely. HMT, when present in the incubation mixture, does not change the hydrolysis pattern (Fig. 3(b)). This proves that no important rearrangement of the secondary structure of 5S RNA occurs after HMT intercalation. Ethidium bromide, when present, induces a cleavage after G37 (Fig. 3(c)). Both drugs unwind DNA to the same degree but ethidium has a much larger association constant (Wiesehahn & Hearst, 1978). The large amount of bound drug must induce an expansion of the 5S RNA which allows better access to G37. Decreasing the ionic strength, even in the absence of Mg^{2+}, does not change the partial hydrolysis pattern. We obtained the same pattern in the buffer used for HMT cross-linking and in the high ionic buffer described by Vigne & Jordan (1977).

The effect of ethidium and HMT on the digestion at low salt (5 mM-NaCl) is the same as that at high salt (results not shown). The presence of HMT has no effect while ethidium induces an additional cleavage after G37.

This is a very strong argument in favor of a stable secondary structure of Drosophila 5S RNA in high and low salt solution. The effects of salt and drugs on the tertiary structure are unknown.

The same crosslinks (XL1-4; Fig. 6) are produced in different salt conditions (Figure 4) and over a wide range of HMT incorporation ratios (data not shown). The fact that each modified fragment is found even at levels of much less than one HMT per 5S molecule shows that they are not the result of induced structure. Once this was established, it was possible to use larger amounts of drug to obtain the quantities needed for analysis. The fact that more than 50% of some bases can be modified under conditions of heavy incorporation suggests that the crosslinking is not occurring in just a small part of the population of conformations, but is, in fact, occurring in the principal species in solution.

The presence of a crosslink does not prove that the involved regions are base-paired in the normal Watson-Crick sense. It does, however, show that the regions are very close because the HMT is very small. Studies on the structure of adducts produced in DNA (Straub et al., 1981) and in RNA (chapter 2) have shown that only one orientation of the HMT in the helix leads to reaction. For a crosslink to occur, the base on the opposite strand must also be in the proper orientation. Stacking or base-pairing in secondary and some types of tertiary structure would allow the correct orientation.

(c) Secondary structure of 5S RNA

Two crosslinks have been demonstrated unequivocally in Drosophila 5S RNA, i.e. crosslinks between fragments 76-82 and 94-97 (XL1; Fig. 6) and between fragments 22-30 and 49-56 (XL4A; Fig. 6). Even in the absence of results from other techniques, these crosslinks, in conjunction with the rules generated for predicting the strength of base pairs (Tinoco et al., 1973; Borer et al., 1974), lead to the secondary structure model shown in Figure 9. This model is identical to that initially proposed for 5S RNA from Torulopsis utilis by Nishikawa & Takemura (1974) and contains elements of the structure proposed by Fox & Woese (1975) and Vigne & Jordan (1977).

Close examination of this model gives new insight into the crosslinks XL2 and XL3. In XL2, the crosslink occurs between a hexanucleotide and a heptanucleotide. There is a heptamer (fragment 42-48) opposite the hexamer 32-37 in the model shown in Figure 9. A crosslink involving C44 or C46 with U33 can be postulated. The low yield of this crosslink would be explained by the weaker reactivity of cytidine (chapter 2). Because the composition of these fragments was not determined, it is possible that an unreversed HMT could make the fragments appear longer. The low yield also makes it impossible to rule out partial digestion of one or both fragments. The position of the intact crosslink and the $^{32}P/^{3}H$ ratio (Table 2) indicate the length must be about 13 bases.

Drosophila melanogaster

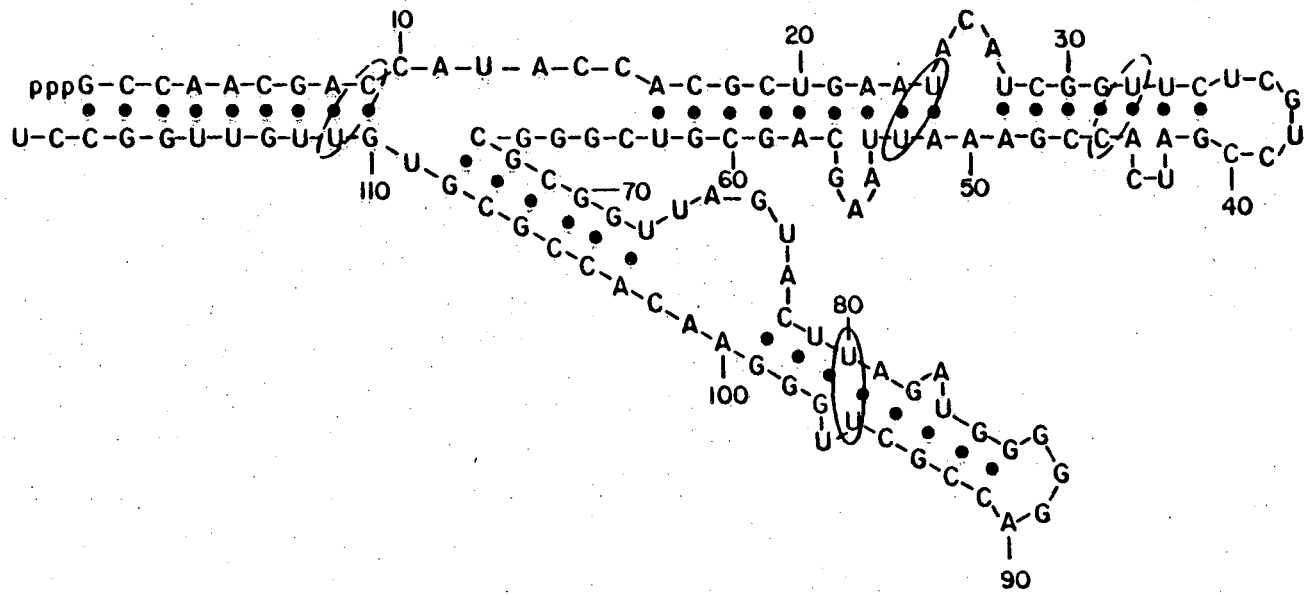


Figure 9: Proposed secondary structure of Drosophila melanogaster 5S RNA. The solid circles show the positions of the crosslinks identified conclusively (XL1 and XL4) while the dashed circles show the crosslinks identified tentatively (XL3 and XL2).

Upon reversal, XL3 produces a fragment 11 bases long. From the position of XL3 in the gel, we can postulate that XL3 either results from the photoaddition of two or three [^{3}H]HMT molecules to a oligonucleotide 11 bases long (a product of partial digestion, for example) or is a product of a crosslink between the fragment 8-18 and a tri- or tetramer. This latter oligonucleotide, due to the very low yield of the crosslink, would not be seen in the analysis. The data shown in Table 2 favor the crosslink hypothesis: only one [^{3}H]HMT photoadduct is present in XL3 because of the $^{32}P/^3H$ ratio. The model we propose in Figure 9 helps in predicting a crosslink with the trimer 111-113. Once again, this would be a U-C crosslink, accounting for the low yield observed.

It can be argued that the crosslinks we were able to detect do not necessarily occur in 5S RNA molecules sharing the same secondary structure. However, the fact that the four crosslinks can easily be integrated into a single model strongly supports the idea that they belong to only one conformation.

It is possible to propose minor variations to the model shown in Figure 9. We have paired bases 33 and 34 with 41 and 42. This decreases the size of the loop to a more favorable number but also introduces a bulge. The rules for predicting secondary structure seem to favor this structure but are not sufficiently refined to answer this question definitely. There are experimental obser-

vations which bear on this point. G41 is mildly reactive to kethoxal in T. utilus (Nishikawa & Takemura, 1978). The low yield might be a result of minor contamination by a partially denatured form or it may merely be an indication of the weakness of the suggested base-pairing. Partial digestion of yeast and HeLa 5S RNA with T_2 RNAse have produced cleavages after G41, but these cleavages occur only after a cut has been made at G37 to open up the loop. Interactions with other components in the ribosome, as suggested by studies with E. coli 5S RNA (Larrinua & Delihas, 1979), may alter the conformation of this region in vivo. Obviously, more experimental evidence is needed. It has been suggested that bases 14, 15 can pair with 64, 65. This is strongly supported by phylogenetic comparisons (Peattie et al., 1981).

It is also possible to draw a different structure for the area around crosslink 1. Because we only determined the T_1 fragments being base-paired and not the specific bases, an alternative scheme which would crosslink U80 with U96 is possible. In this case, A81 and G82 would be paired to U96 and U95. Although this would replace a GC pair with a GU pair, two bulges would be replaced by a single internal loop. As shown in Table 3, the stabilities are comparable. We have chosen the pairing scheme shown in Figure 9 because it agrees better with the partial enzymatic digestion data and also because the same type of structure is more stable in HeLa 5S RNA.

The secondary structure proposed here is easily generalized to other eukaryotes (Saccharomyces cerevisiae for example, Fig. 10). The stabilities of the various regions for S. cerevisiae and HeLa 5S RNA are given in Table 3. It is also possible to generalize the model to prokaryotes (Fox & Woese, 1975; Vigne & Jordan, 1977). In this case, the helix which pairs the regions around 70 and 105 is present but its stability varies considerably. In thermophilic bacteria, the base pairing is very strong, while in more temperate species, the base pairing is weak (Stahl et al., 1981).

(c) Validity of the model

Any proposed secondary structure must, of course, be compatible with all the available information. Most of the studies on 5S RNA have dealt with E. coli. Because of the differences between prokaryotic and eukaryotic 5S RNA it is sometimes difficult to use data concerning E. coli 5S RNA to discuss the validity of models proposed for eukaryotic 5S RNA. The following data, most dealing with eukaryotic 5S RNA, argue in favor of the model presented here (Fig. 9).

The model presented in Figure 9 differs from the evolutionary model in two respects. The region around base 90 has been paired. This structure can be drawn for all eukaryotes. The additional base-pairs between bases 22-24 and 51-53 are present in some, but not all, eukaryotes. In species which do not have these additional base-pairs,

Saccharomyces cerevisiae

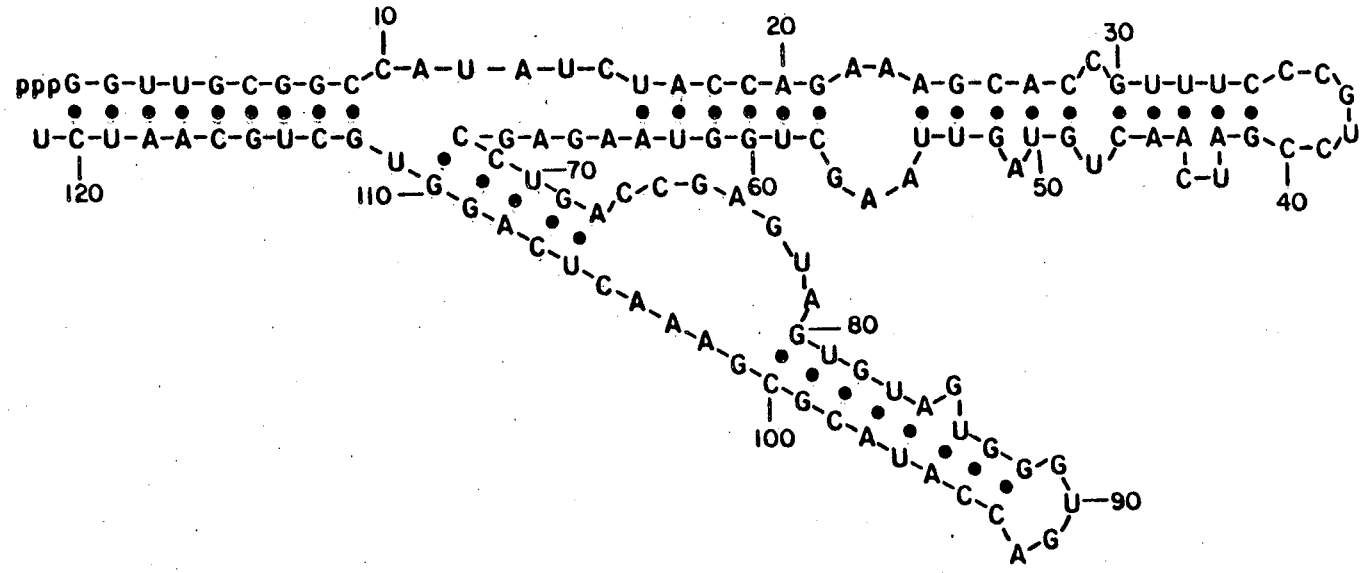


Figure 10: Proposed secondary structure of Saccharomyces cerevisiae 5S RNA.

the adjoining helices are more stable.

Some of the most reliable data come from the partial ribonuclease digestions. The first cleavages in most 5S RNAs are usually around positions 40 and 90. The reasons for this are obvious after examining the secondary structure (Fig. 9). The hairpin loop around base 90 is one of the two most accessible regions and the small size of the loop causes it to be the most strained part of the molecule, so it is understandable that the first cut is made here. The positions of 19 cleavages have been examined for Drosophila 5S RNA (Benhamou et al., 1977). One of these, after G85, occurs in conjunction with the cut G89. Once the cut at G89 has been made, the neighboring helix would no longer be stable and could easily be chewed away by T_1. In fact, very mild digestion with T_1 shows that the first cut is made after G88 with lesser cutting occurring at G87 and G89. No initial cutting is made after G85 (Fig. 11). Of the remaining 17 cleavages, only three are in regions that are not predicted to be either single-stranded or bulged. These three, after G18, U80 and G110, are made only after extensive cleavages in other parts of the molecule. In fact, the enzymatic cuts are remarkably good at predicting the ends of helices and bulges. More recent work in which care was taken to examine only initial cutting (Douthwaite & Garrett, 1981; Troutt et al., 1982), shows that all of the primary cuts agree with this secondary structure. In addition, studies using cobra venom ribonuclease (double

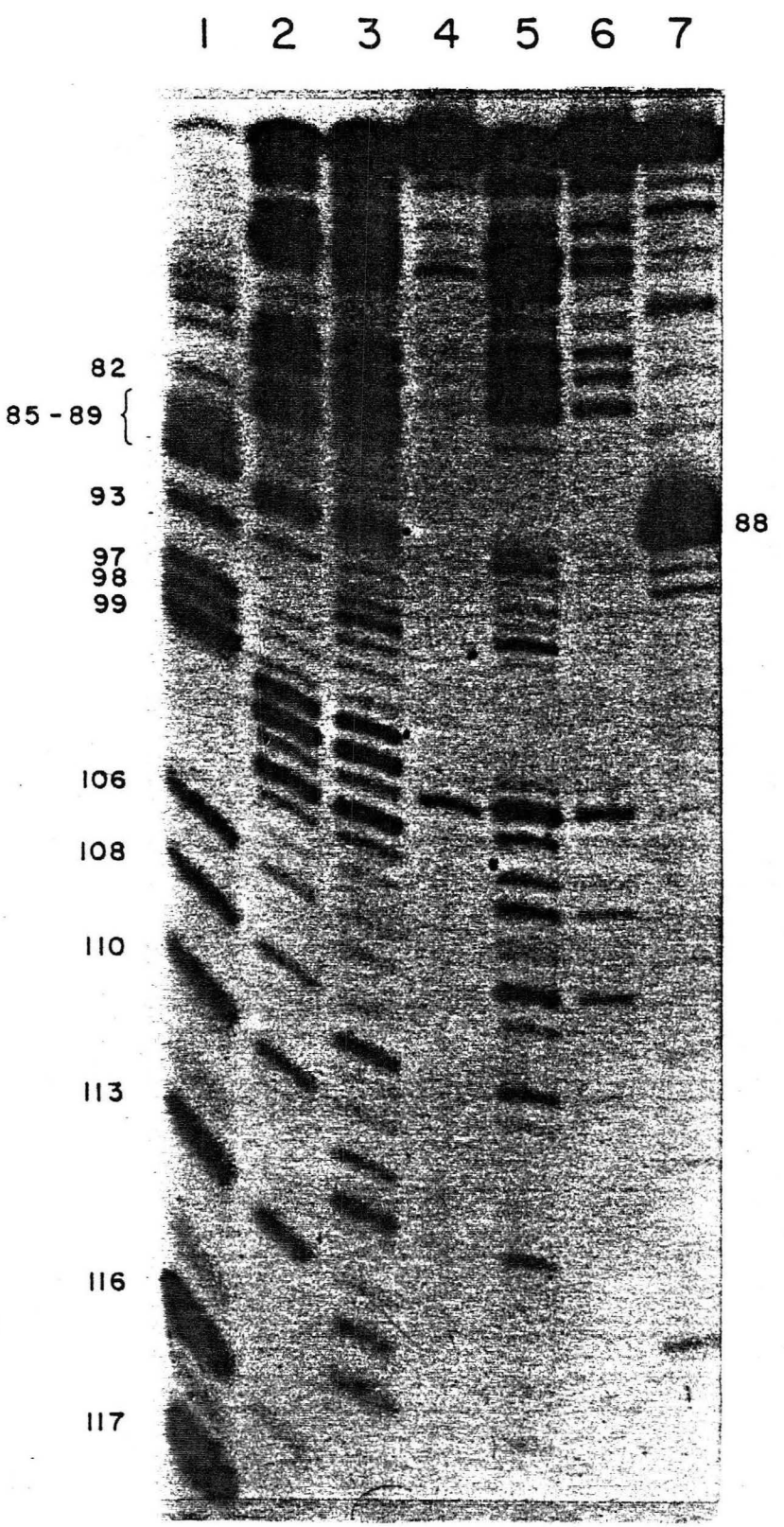

1 2 3 4 5 6 7
82
85 - 89
93
97
98
99
106
108
110
113
116
117
88

Figure 11: Partial enzymatic digestion of 5S RNA. 3'[^{32}P] 5S RNA was digested under mild conditions such that each molecule would have one or less cleavages. Lanes 1. 2, 3, and 5 were done under denaturing conditions so a sequence ladder was generated. Reactions were done as suggested by the supplier (PL Biochemicals). Lane 1 contains bands corresponding to G (T_1 RNase), lane 2 contains A (U_2 RNase), lane 3 contains A and U (Phy M RNase) and lane 5 has U and C (B. cereus RNase). Lanes 4 and 6 contained 10 ng/ml pancreatic RNase, 0.1 M Tris-HCl (pH 7.5), 10 mM EDTA and were digested for 30 min at 0° and 37°, respectively. Lane 7 contained 20 units/ml T_1 RNase, 0.1 M Tris-HCl (pH 7.5), 10 mM EDTA and was digested for 30 min at 0°. The principal cut in lane 7 is after G88 while the principal resolved cut in lane 4 is after C102. Both are predicted to be single-stranded in Figure 10. Numbering of the the G's is shown adjacent to lane 1 for convenience.

strand specific) also support this model.

Chemical modification of exposed guanines with kethoxal has been done for T. utilis 5S RNA (Nishikawa & Takemura, 1978). All of the strongly reacting sites fell within single-stranded regions predicted by our model. Laser raman studies on S. cerevisiae 5S RNA have predicted that 65% of the uridines are involved in base-pairing with a total of at least 35 base-pairs (Luoma & Marshall, 1978a). When the crosslinking model is generalized to yeast, 68% of the uridines and 38 base-pairs are observed (Fig. 10). Infrared spectroscopy (Stulz et al., 1981) and NMR (Luoma et al., 1980) also indicate there are 35-40 base pairs in this 5S molecule.

Three crosslinking studies have been done with E. coli 5S RNA. Rabin and Crothers (1979) found a crosslink in the stem region using a psoralen derivative, AMT. Wagner and Garrett (1978) found a similar crosslink using 1,4 phenyldiglyoxal. As mentioned earlier, the glyoxal functionalities are specific for single stranded guanines while the phenyl group promotes intercalation into double stranded regions. Further work with this reagent has resulted in two additional crosslinks (Hancock & Wagner, 1982). One of these was arbitrarily chosen to reflect true structure while the other was dismissed as caused by structural heterogeneity. While a new secondary structure was proposed on the basis of this crosslink, a more proper conclusion would have been that this reagent is not suitable for secondary

structure analysis. It appears to unfold the 5S and cause normally protected guanosines to react with a probe which is inherently specific for single stranded regions.

(d) Stability of the model

Two different methods for predicting the stability of secondary structures have been used to estimate the stability of the model (Table 3). Both of these methods have their drawbacks when used to analyze a large molecule (relative to the model compounds used) like 5S RNA with unknown tertiary structure. The values obtained are a first approximation of the stability. Unlike other models, the model proposed here has approximately the same stability among all eukaryotic species (Table 3). All of the regions are also independently stable.

Melting studies have been done on S. cerevisiae 5S RNA (Maruyama et al., 1979; Luoma et al., 1980) but the results do not agree with each other. Maruyama et al. (1979) found that all helices melted cooperatively. Since the 3 regions in this model are all of approximately the same stability, this can be understood. Other models do not predict this behavior. Luoma et al. (1980) found a biphasic melting with half the bases melting prior to the other half. The uncertainty in the stability calculations also allow this possibility. If the hairpin 40 arm (the most unstable) melted first, this behavior would be predicted. What is clear, is the need for careful studies in a variety of salt conditions which would remove this ambiguity.

Table III: Stability of Drosophila melanogaster, Hela, and Saccharomyces cerevisiae 5S RNA

		Stability (in kcal/mole)					
		Thermodynamic[1]			Empirical[2]		
Model	Region	D.m.	Hela	S.c.	D.m.	Hela	S.c.
This work	Stem	18.2	15.5	16.1	18.9	15.9	15.5
	Hairpin 40 arm	20.0	19.0	10.3	20.7	18.3	13.2
	Hairpin 90 arm	11.6	16.4	15.3	5.9	5.7	15.3
	(Alternate[3])	(12.5)	(14.6)		(5.9)	(4.3)	
	Total	49.8	50.9	41.7	41.4	36.2	40.3
Vigne and Jordan (1977)	Stem	18.2	15.5	16.1	18.9	15.9	15.5
	Hairpin 40 arm	14.6	14.8	5.2	18.5	16.4	8.8
	Hairpin 90 arm	6.2	6.3	3.1	4.0	3.2	3.7
	Total	39.0	36.6	24.4	37.3	31.8	24.3
Luoma and Marshall (1978a)	Stem	24.7	24.3	17.9	22.7	21.2	16.9
	Hairpin 40 arm	1.8	7.0	+3.9	2.1	3.7	+0.1
	Hairpin 65 arm	1.3	6.0	1.1	+2.4	7.6	2.5
	Hairpin 90 arm	0.5	6.1	7.2	+2.2	+1.9	6.4
	Total	28.3	43.4	22.3	16.7	27.5	20.4

[1]Calculated from data given by Tinoco et al. (1973), Borer et al. (1974), and Gralla and Crothers (1973).

[2]Calculated from data given by Ninio (1979).

[3]See text for description of alternate base-pairing scheme for the region around base 90.

Small angle X-ray scattering measurements have provided a model for the gross shape of 5S RNA (Osterberg et al., 1976; Muller et al., 1981). An elongated, Y-shaped molecule with an axial ratio of 5:1 has been predicted. This shape can be generated in a number of ways with the model presented in Figure 9. Some idea of the types of tertiary interactions, if any, which are occurring is needed before a three-dimensional model can be formulated with any degree of confidence.

(f) 5S RNA in ribosome

It is possible to overinterpret the results of studies on 5S RNA in solution when theorizing on structure in the ribosome. Unlike tRNA, 5S RNA does not normally function free in solution. Studies of 5S RNA in the ribosome are more difficult because much of the molecule is not in contact with solution. The high salt needed to maintain ribosomal structure and the large number of proteins both serve to inhibit intercalation of HMT. Absolutely no crosslinking of 5S RNA in the intact ribosome could be detected.

REFERENCES

Aubert, M., Scott, J. F., Reynier, M., and Monier, R. (1968) Proc. Nat. Acad. Sci. USA 61, 292-299.

Benhamou, J., Jourdan, R., and Jordan, B. R. (1977) J. Mol. Evol. 9, 279-298.

Borer, P. N., Dengler, B., Tinoco, I. Jr., and Uhlenbeck, O. C. (1974) J. Mol. Biol. 86, 843-853.

Brown, D. D. and Littna, E. (1964) J. Mol. Biol. 8, 669-687.

Dohme, F. and Nierhaus, K. H. (1976) Proc. Nat. Acad. Sci. USA 73, 2221-2225.

Douthwaite, S. and Garrett, R. A. (1981) Biochemistry 20, 7301-7307.

Echalier, G. and Ohanessian, A. (1970) In vitro 6, 162-172.

Erdmann, V. A. (1976) Progr. Nucl. Acid Res. Mol. Biol. 18, 45-90.

Erdmann, V. A. (1980) Nucl. Acids Res. 8, r31-r47.

Fox, G. E. and Woese, C. R. (1975) Nature 256, 505-507.

Gralla, J. and Crothers, D. M. (1973) J. Mol. Biol. 73, 497-511.

Hancock, J. and Wagner, R. (1982) Nucl. Acids Res. 10, 1257-1269.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. and Zamir, A. (1965) Science 147,1462-1465.

Hyde, J. E. and Hearst, J. E. (1978) Biochemistry 17, 1251-1257.

Jordan, B. R., Jourdan, R. and Jacq, B. (1976) J. Mol. Biol.

101, 85-105.

Larrinua, I. and Delihas, N. (1979) Proc. Nat. Acad. Sci. USA 76, 4400-4404.

Luoma, G. A. and Marshall, A. G. (1978a) J. Mol. Biol. 125, 95-105.

Luoma, G. A. and Marshall, A. G. (1978b) Proc. Nat. Acad. Sci. USA 75, 4901-4905.

Luoma, G. A., Burns, P. D., Bruce, R. E. and Marshall, A. G. (1980) Biochemistry 19, 5456-5462.

Maruyama, S., Tatsuki, T., and Sugai, S. (1979) J. Biochem. 86, 1487-1494.

Monier, R. (1974) in Ribosomes (Nomura, M. et al., eds.) pp.141-168, Cold Spring Harbor Laboratory, New York.

Muller, J. J., Welfle, H., Damaschun, G., and Bielka, H. (1981) Biochim. Biophys. Acta 654, 156-165.

Musajo, L., Bordin, F., Caporale, G., Marciani, S. and Rigatti, G. (1967) Photochem. Photobiol. 6, 711-719.

Ninio, J. (1979) Biochimie 61, 1133-1150.

Nishikawa, K. and Takemura, S. (1974) FEBS Letters 40, 106-109.

Nishikawa, K. and Takemura, S. (1978) J. Biochem. 84, 259-266.

Osterberg, R., Sjoberg, B., and Garrett, R. A. (1976) Eur. J. Biochem. 68, 481-487.

Pace, B., Matthews, E. A., Johnson, K. D., Cantor, C. R., and Pace, N. R. (1982) Proc. Nat. Acad. Sci. USA 79, 36-40.

Peattie, D. A., Douthwaite, S., Garrett, R. A., and Noller,

H. F. (1981) Proc. Nat. Acad. Sci. USA 78, 7331-7335.

Rabin, D. and Crothers, D. M. (1979) Nucl. Acids Res. 7, 689-703.

Rosset, R. and Monier, R. (1963) Biochim. Biophys. Acta 68, 653-656.

Sommer, S. (1979) Anal. Biochem. 98, 8-12.

Stahl, D. A., Luehrsen, K. R., Woese, C. R., and Pace, N. R. (1981) Nucl. Acids Res. 9, 6129-6137.

Straub, K., Kanne, D., Hearst, J. E. and Rapoport, H. (1981) J. Am. Chem. Soc. 103,2347-2355.

Stulz, J., Ackermann, T., Appel, B., and Erdmann, V. A. (1981) Nucl. Acids Res. 9, 3851-3861.

Tinoco, I. Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973) Nature New Biol. 246, 40-41.

Troutt, A., Savin, T. J., Curtiss, W. C. Celantano, J., and Vournakis, J. N. (1982) Nucl. Acids Res. 10, 653-664.

Vigne, R. and Jordan, B. R. (1977) J. Mol. Evol. 10, 77-86.

Wagner, R. and Garrett, R. A. (1978) Nucl. Acids Res. 5, 4065-4075.

Wegnez, M., Denis, H., Maabraud, A., and Clerot, J. C. (1978) Develop. Biol. 62, 99-111.

Wiesehahn, G. and Hearst, J. E. (1978) Proc. Nat. Acad. Sci. USA 75, 2703-2707.

Chapter Six:

Function and Structure of *E. coli* 16S RNA by Psoralen Crosslinking

INTRODUCTION

The secondary structure of *E. coli* 16S RNA has been speculated on ever since the first, partial sequences were determined (Fellner et al., 1970). Real progress was not made until the complete sequence of the RNA was unequivocally determined (Brosius et al., 1978; Carbon et al., 1979). Once the sequence was established, chemical modification and enzymatic digestion data (reviewed by Noller and Woese, 1981) could be used with greater confidence. Availability of the small subunit RNA sequences of *Z. mays* chloroplasts (Schwarz & Kossel, 1980), *P. vulgaris* (Carbon et al., 1981), *S. cerevisiae* (Ruptsov et al., 1980), *X. laevis* (Salim & Maden, 1981) and various mitochondria (Eperon et al., 1980; Van Etten et al., 1980; Sor & Fukuhara, 1980) has made phylogenetic comparisons possible. On the basis of this data, Noller & Woese (1981) and Stiegler et al., (1981) have proposed a general secondary structure model for all small subunit RNAs. Zwieb et al. (1981) have also proposed a general model based on this evidence as well as UV crosslinking (Zwieb & Brimacombe, 1980) and denaturation studies (Ross & Brimacombe, 1979; Glotz & Brimacombe, 1980). All three models are very similar.

Despite this agreement, the models are far from com-

plete. Psoralen crosslinking data generated by electron microscopy yields results that, in many instances, cannot be easily incorporated into the other models (Wollenzein et al. 1979; Wollenzein & Cantor, 1982). Because of the low resolution of the electron microscope, the unambiguous placement of the crosslinks is not possible. It is clear that psoralen crosslinking is generating new information, so this study has been undertaken to obtain more detailed localization of crosslinks and thus a more complete picture of the 16S RNA secondary structure.

A technique, based on that of Zwieb & Brimacombe (1980) has been developed which allows resolution to at least ± 15 bases. Because psoralen reacts with only a certain class of sites and only a few secondary structures can be drawn for any given fragments, actual resolution is generally to the exact nucleotides crosslinked.

The function of certain regions in 16S RNA is already known. Bases 1393-1401 are involved in binding tRNA to the P site (Taylor et al., 1981) and bases 1534-1540 bind mRNA (Shine & Dalgarno, 1975; Steitz & Jakes, 1975). Other regions have been implicated in other functions. Correlation of the data generated by psoralen crosslinking with other results enables us to formulate detailed mechanisms for the way in which ribosomes carry out protein synthesis. While some aspects of these mechanisms are purely speculative, predictions made can be tested to gain further insight into ribosomal functioning.

MATERIALS AND METHODS

E. coli MRE 600 cells were grown as by Traub et al. (1971). Frozen cells (1 g) were suspended in 10 ml of 50 mM sodium acetate (pH 5.0) and 10 mM VRC (vadanylribonucleoside complex; Berger & Berkenmeier, 1979). Cells were homogenized in a ground glass tissue homogenizer until the solution became viscous. DNase (Worthington, RNase free) was added to about 50 µg/ml and homogenization continued until the solution was no longer viscous. An equal volume of redistilled phenol was added. Phenol extraction was repeated at least three more times (until the water-phenol interface was clear). The solution was made .2 M NaCl and ethanol precipitated at -20° twice. The precipitated RNA was dissolved in a minimum volume of .1 M LiCl, 10 mM EDTA, 0.5% SDS, 10 mM Tris base (LES). Approximately 5 mg was placed on top of a 15-30% sucrose gradient containing LES buffer. The RNA was centrifuged for 24 hr at 27,000 rpm in an SW27.1 rotor. Bands corresponding to 23S RNA and 16S RNA were resolved by pumping the solution through a Beckman Analytical Optical Unit (254 nm). 16S RNA fractions were combined according to purity and precipitated. When necessary, the centrifugation was repeated in order to get pure 16S RNA. The 5S/tRNA band was sometimes obscured by residual VRC.

Purity and intactness of the RNA was examined by gel electrophoresis in 4% polyacrylamide (20:1, w/w, acrylamide to bis-acrylamide), 7 M urea gels buffered with 50 mM Tris-borate (pH 8.3), and 10 mM EDTA. Polymerization was cata-

lyzed by .075 g ammonium persulfate and 50 µl TEMED per 100 ml. All gel materials were purchased from Bio Rad except urea which was from Schwarz-Mann.

Before crosslinking, 16S RNA was incubated for 30 min or more at 37° in TMA I buffer (10 mM $MgCl_2$, 100 mM NH_4Cl, 10 mM Tris-HCl, pH 7.2, 14 mM β-mercaptoethanol). Crosslinking was done in the apparatus described previously (Thompson et al., 1981) at 10°. Three additions of HMT were made from a stock solution in DMSO (2 mg/ml). Each addition made the aqueous solution 20 µg/ml in HMT and additions were separated by 10 min. This protocol produced an average of 3-5 HMTs per 16S molecule. After phenol extraction and ethanol precipitation, the RNA was redissolved in 50 mM Tris-HCl (pH 8.5), 10 mM $MgCl_2$ at a concentration of 10 OD_{260}/ml. This was digested for 2 hr at 37° with 100 u/ml T_1 RNase (Sigma). This was phenol extracted and ethanol precipitated.

The digested RNA was then separated by two different methods. Method one involved running the RNA through an RPC-5 column (0.4 x 20 cm; Astro Enterprises) and eluting with 80 ml of a 0.1 to 2 M KCl gradient. The KCl solutions also contained 2 mM sodium thiosulfate, 10 mM EDTA, .06% sodium azide, and 10 mM Tris-HCl (pH 6.8). One ml fractions were collected and ethanol precipitated. Later fractions required dialysis to remove excess KCl before precipitation. Method two involved running the RNA through a BND cellulose column (0.5 ml, Sigma) with a 15 ml 0% DMSO, 0.3 M

NaCl to 25% DMSO, 0.65 M NaCl gradient for elution.

The purified, precipitated RNA was redissolved in 25 µl of 50 mM Tris-HCl (pH 8.5), 10 mM $MgCl_2$, 13 mM β-mercaptoethanol, 2 mM spermine and labelled with 0.5 mCi of [γ-^{32}P] ATP (ICN, crude) and 2 units of polynucleotide kinase overnight at 37°. The sample was then made 1 M in urea, heated at 90-95° for 1 min, fast cooled on ice and loaded onto a .08 x 15 x 40 cm 12% polyacrylamide gel (made like the 4% gel described earlier except no urea was present). This gel was run at room temperature until the bromophenol blue had migrated off the gel (about 7 hr at 800 V). The upper glass plate was removed and the gel covered with plastic wrap. The lane(s) containing the RNA was cut using a razor blade and a straight edge. The lower 30 cm was placed on a 34 x 43 cm plate. After removal of the plastic wrap, a 20% polyacrylamide (30:1, acrylamide to bis), 7 M urea gel was polymerized around the gel strip. The gel was buffered at twice the salt concentration of the first dimension gel. The second dimension was run warm until the xylene cyanol marker had migrated off the gel. The gel was covered with plastic wrap and autoradiographed with Kodak XAR-5 film. Crosslinks were reversed in the gel by exposure to a 40 W germicidal lamp at a distance of about 10 cm for 2 hr. After reversal, fragments of interest were cut from the the gel and eluted in 200 µl of LES buffer for 6-16 hours twice. Ten micrograms of tRNA was added and the samples ethanol precipitated The RNA was run on another 20% polyacrylamide, 7 M urea gel

to separate the previously crosslinked fragments. After autoradiography, the fragments of interest were eluted from the gel as described above.

Fragments were sequenced enzymatically. Digestions were carried out at 55° for 15 min in 10 µl which contained 20 mM sodium citrate, 1 mM EDTA, and 2 µg carrier tRNA. In addition, the RNase T_1 (G specific), U_2 (A specific), and Phy M (A and U specific) contained 7 M urea while the B. cereus (U and C specific) did not. The U_2 reaction was buffered at pH 3.5 and the others at pH 5.0. Phy M and B. cereus were purchased from P. L. Biochemicals and used as described. T_1 and U_2 were purchased from Sigma and dissolved in stock solutions at 50 u/ml. One microliter of the stock solution was needed for each digestion. After reaction, the samples were run on 20% polyacrylamide gels and autoradiographed using Dupont Lightening Plus intensifying screens.

RESULTS

A schematic of the methodology used to form and analyze crosslinks is shown in Figure 1. Crosslinking was done in reconstitution buffer at low levels of HMT reaction in order to attain the most biologically relevant conformation. HMT has been shown to have a minimal effect on RNA structure when added at these low doses, 3-5 HMTs per 16S (Thompson et al., 1981; Thompson et al., 1982). The crosslinked and uncrosslinked 16S melted at exactly the same temperature in reconstitution buffer. The psoralen sample had about 10%

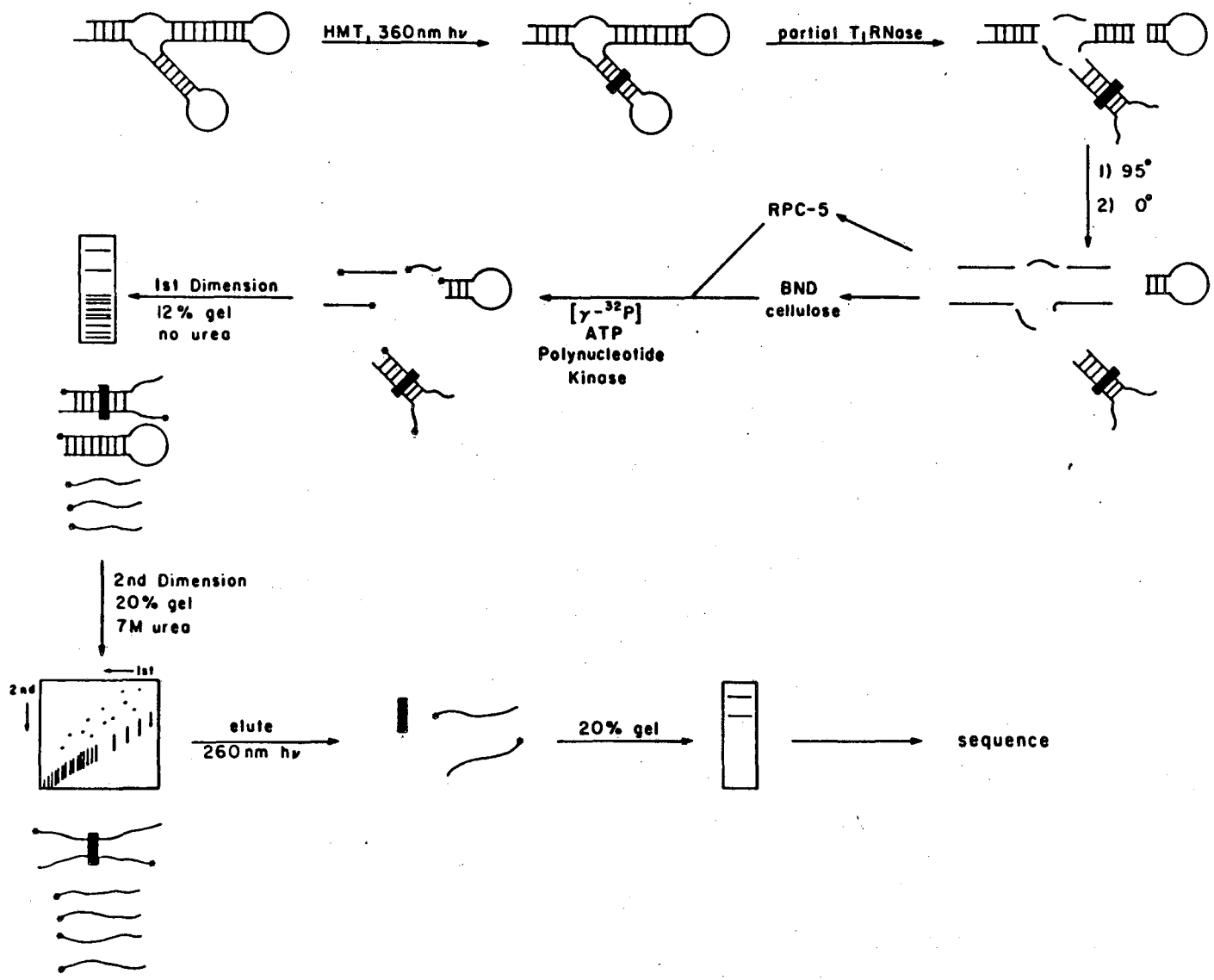

HMT, 360nm hν
partial T1RNase
1) 95°
2) 0°
RPC-5
BND cellulose
[γ-32P] ATP Polynucleotide Kinase
1st Dimension
12% gel
no urea
2nd Dimension
20% gel
7M urea
1st
2nd
elute
260nm hν
20% gel
sequence

Figure 1: Schematic diagram of methods used to analyze crosslinks.

less hyperchromicity and also renatured much more quickly (results not shown). When examined at lower ionic strength, the crosslinked sample melted at lower temperatures, suggesting that a different structure was more stable. The crosslinked sample was locked in the reconstitution conformation and could not change as easily as the uncrosslinked sample.

Zwieb and Brimacombe (1980) have used a similar two dimensional gel technique to examine UV induced crosslinks. In the first dimension gel, any pre-existing secondary structure is stable because there are no denaturants and the gels are run at room temperature. The only fragments which will have structure are snapback hairpins and covalent cross links. Prior to loading on the gel, all intermolecular secondary structure is destroyed by heating and rapid cooling. This produces a more uniform population of molecules. The second dimension contains molecules with no secondary structure because it contains 7 M urea and is run hot. Most molecules are unaffected because they contained no secondary structure in the first dimension and thus run as a diagonal. Hairpin loops actually run slightly faster in the second dimension because their radius decreases somewhat. Covalently crosslinked molecules, however, are severely retarded in the second dimension. After melting, crosslinks have four ends an are forced to move through the gel in a spread out, octopus-like conformation (Zwieb and Brimacombe, 1980).

As shown in Figure 2, the only bands which appear above the diagonal are caused by crosslinking. In 2D gels which

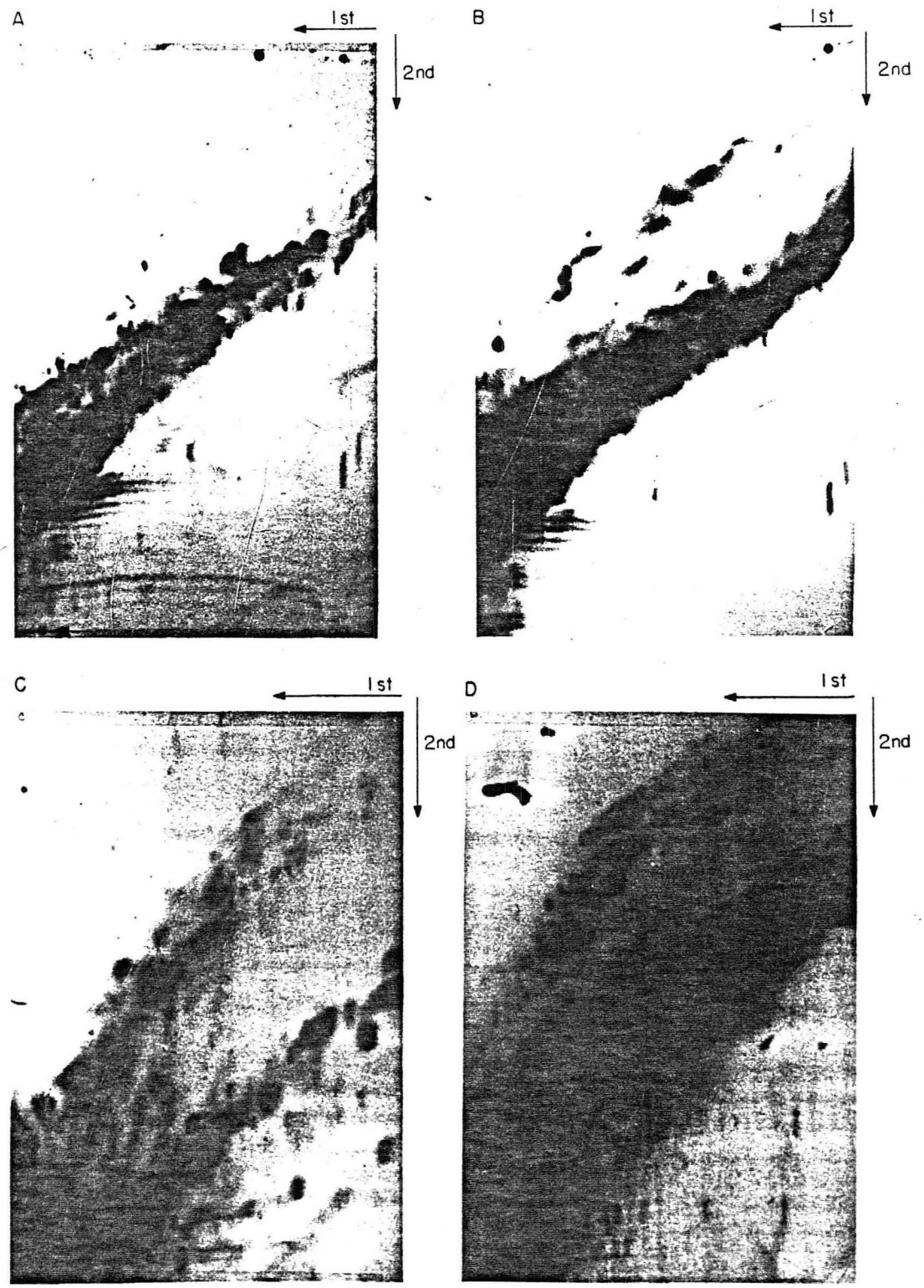

A
1st
2nd
B
1st
2nd
C
1st
2nd
D
1st
2nd

Figure 2: Autoradiograms of two-dimensional gels. A) 2D gel of uncrosslinked 16S RNA. Heavy T_1 digestion results in very little large material near the top of the gel. The "X" marks the position of xylene cyanol. B) 2D gel of crosslinked 16S RNA. This sample was treated exactly as that in A except it was crosslinked as described in Materials and Methods. Discrete off diagonal spots result from heavy T_1 digestion. C) 2D gel of fraction from RPC-5 column. Note the two diagonals, both of which are heavier near the bottom of the gel. D) 2D gel of fraction from BND cellulose column The family of dark, off diagonal spots near the center are all GPs 1116 x 1183. Other fractions yield gels with a much different pattern of off diagonal spots.

contain total 16S RNA, the number of off-diagonal spots is simply too large to handle. Only a few crosslinks can be retrieved in pure form with enough ^{32}P to sequence. In order to circumvent both these problems, the T_1 digests of crosslinked RNA was first run through either a RPC-5 or BND cellulose column. Fractions from these columns were then labelled and analyzed by the 2D gel system. This protocol has the double advantage of reducing cross-contamination of crosslinked bands and also providing better incorporation of [γ-^{32}P]ATP.

The basis for RPC-5 separation of crosslinks is not entirely clear. What is observed is the presence of two diagonals. The lower diagonal corresponds to the normal single stranded material while the upper, steeper diagonal consists entirely of crosslinked material. When a 0.1 to 2.0 M KCl gradient was used for elution, the most useful fractions for analysis came off the column at .45-.80 M KCl. Later fractions tended to contain material too large for analysis. This can be recombined, redigested, and run on the column again.

BND cellulose separates principally on the basis of secondary structure (Sedat et al., 1967). Before adding RNA to the column, it was heated and quick cooled to minimize non-crosslinked secondary structure. The profile of a typical column eluent is shown in Figure 3. It is quite clear that the OD and ^{3}H (from ^{3}H-HMT) peaks do not coincide. This is as expected. Most RNA has little secondary struc-

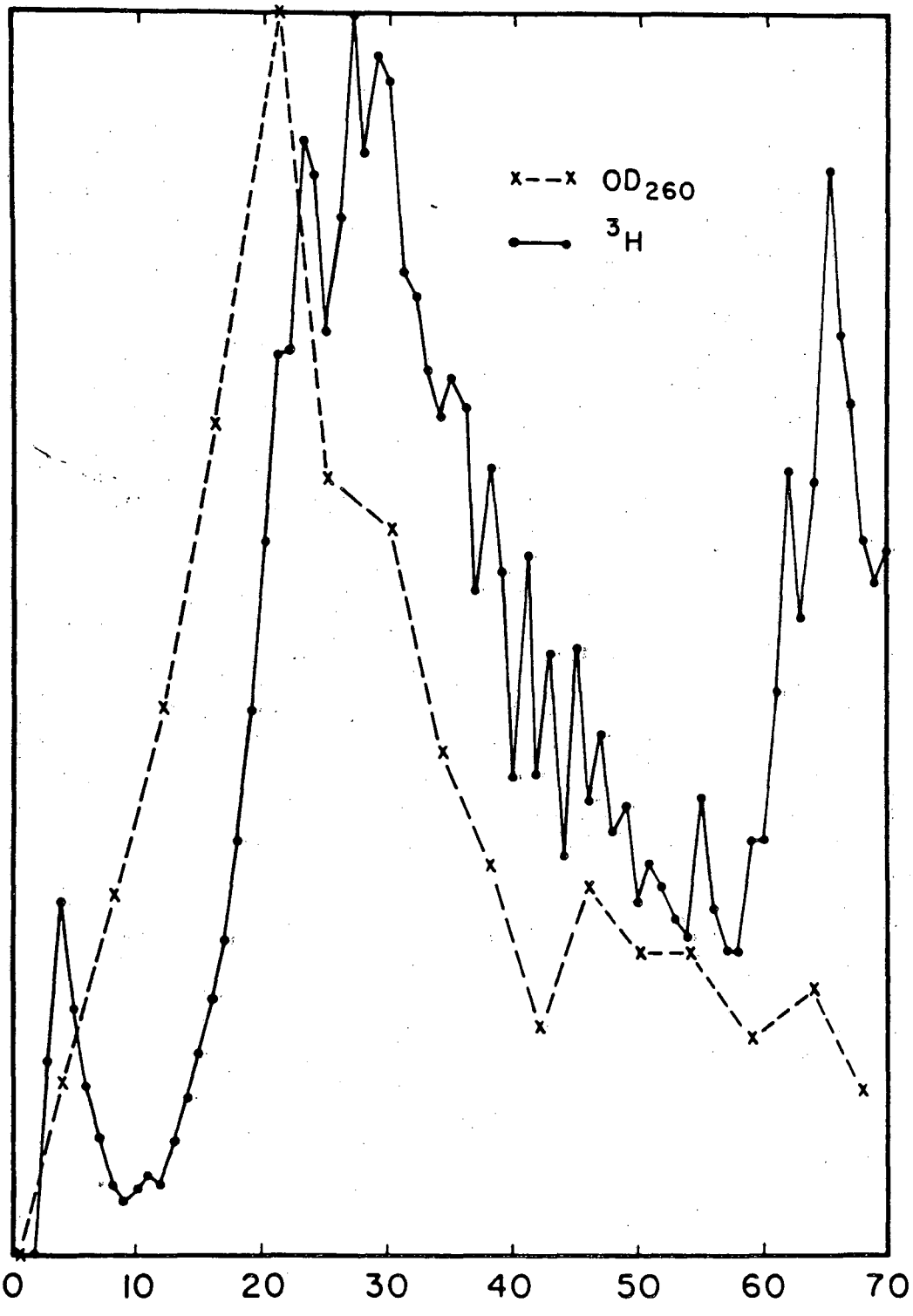

x--x OD260
3H
0
10
20
30
40
50
60
70
Fraction

Figure 3: Elution profile of BND cellulose column. ^{3}H arising from [^{3}H]HMT is plotted versus fraction for a BND cellulose column run as described in Materials and Methods (⟷). Relative OD_{260} of selected fractions is also shown (×--×).

ture and elutes early. The later maxima in OD (fractions 45 to 65) probably result from the presence of hairpin loops. The ^3H peak elutes more slowly because the crosslinks stabilize the secondary structure to the denaturing effect of DMSO. The second ^3H peak contained a great deal of crosslinked hairpin loops while the earlier ^3H peak tended to contain more longer range crosslinks with less complementarity. There were significant amounts of hairpin structures also, though. Fractions beyond fraction 70 were also collected but ^3H counts dropped quickly beyond this. This material tended to be very large but could be redigested as before.

The crosslinks obtained varies tremendously with the level of T_1 digestion. If the samples are heavily digested as in Figure 2b (note the lack of material near the top of the diagonal), a small set of crosslinks is found. Light digestion results in more crosslinks which are less well resolved. If the full range of crosslinks is to be found, several different digestions must be done.

In order for the two dimensional gel system described here to separate crosslinked and uncrosslinked material, the fragments must each have a minimum length of approximately 15 nucleotides. Complete digestion with any RNase would make the fragments too small to be of use. In order to produce oligonucleotides that are large enough to separate but small enough to be resolvable, partial digestion was done with T_1 RNase. This has the undesirable property of produc-

ing many different sized fragments which actually contain the same crosslink. The number of off-diagonal spots produced even when using a crosslinking reagent as specific as psoralen becomes overwhelming. To compound this even further, it was observed that many spots can reverse to yield exactly the same pair of fragments. This appears to be primarily a problem in the running of the first dimension gel.

To reduce the number of off-diagonal bands, first dimension gels containing a short 4% polyacrylamide, 7 M urea stacking gel were employed. This ensured that no intermolecular base pairing could be present. While this did reduce the number of bands, it also reduced the separation of bands from the diagonal. The decision of which type of first dimension gel to use is thus dependent upon the seriousness of the multiple spot phenomenon.

After photoreversal of the crosslink and elution from the gel, fragments were separated on a 20% polyacrylamide, 7 M urea gel. In some cases a single band was observed. If the mobility of this band changed distinctly after reversal (R_f in 2nd dimension gel and separation gel were compared), the band was classified as a hairpin loop and sequenced. When two bands were found, both were sequenced. Frequently, more than two bands were found. Unfortunately, relative intensities of labelling cannot be used to match pairs of bands because the halves of crosslinks are not necessarily labelled equally. The 5' terminal bases are usually different and the ends are usually differentially offset in sec-

ondary structure. These ambiguous crosslinks were discarded.

Because the fragments are 5' end labelled, rapid sequencing techniques can be employed in determining their positions in 16S RNA. Many fragments were sequenced simultaneously so the less time consuming method of partial enzymatic digestion was chosen over chemical cleavage. The sequence of 16S RNA is already known (Brosius et al., 1978) so ambiguitiies in enzymatic cuts are easily resolved. The positions of the 5' and 3' bases in the unambiguous crosslinks are shown in Table I. Frequently, different sets of termani were obtained from different off diagonal spots for the same crosslink. These overlapping fragments allow an even finer resolution of the crosslink position.

DISCUSSION

Reactivity of Psoralen

In order to best assign the locations of crosslinks within the observed fragments, a knowledge of the specificity of reaction of HMT is necessary. Reaction of different psoralens with various RNAs (Bachellerie et al., 1981; Thompson et al., 1982) and DNAs (Straub et al., 1981; Kanne et al., 1982) has demonstrated that uridine and thymine are the preferred sites for photoreaction. There are only a few studies in which the position of psoralen has been mapped to high resolution (Rabin and Crothers, 1979; Thompson et al., 1981; Bachellerie and Hearst, 1982; Youvan and Hearst, 1982; Turner et al., 1982). These studies indicate that the most

Table I: Locations of Crosslinked Oligonucleotides

FRAGMENT 1		FRAGMENT 2	
5' Ends	3' Ends	5' Ends	3' Ends
	Confirmatory Crosslinks		
238	289	-	-
435	481	-	-
576, 592	616	617	645, 650, 655
994	1015	1016	1043
1317, 1335	1361	1362	1379, 1385
1280	1316	1317	1361
1234	1255	1280	1304
	Discriminatory Crosslinks		
1100	1127	1167, 1179	1206, 1215
1167, 1179	1206, 1215	-	-
7	46*	918	941
	New Crosslinks		
327	369	1317	1343
618	645	1405, 1418	1432, 1453
955	976	1498	1514

*This croslink oligonucleotide is tentative. See Discussion.

Multiple 5' and 3' ends are found because most crosslinks have been sequenced more than once with different termani present in the different off diagonal spots.

reactive positions occur in locations which are relatively unstable and facilitate intercalation Particularly susceptible sites occur near the ends of helices, adjacent to G-U pairs, or at the end of runs of uridine. Certainly, there are non-Watson-Crick pairing schemes which could be crosslinked. Such structures may be inherently less reactive because of the rigidity which they impose on nearby helices. In tRNA, the melting of the tertiary structure greatly enhanced psoralen reaction (Bachellerie and Hearst, 1982).

An entirely different type of tertiary interaction has been observed in tRNA (Kim et al., 1974). Coaxial stacking of helices appears to be a major structural feature in tRNA and has also been proposed as an important feature of large ribosomal RNAs (Noller et al., 1981). Such structures may be reactive if they can be unwound sufficiently to allow stacking of the psoralen between helices. The non-continuous phosphodiester backbone should allow unwinding to occur more easily.

Nomenclature

The nomenclature used to describe the crosslinks found is an extension of that used by Noller et al. (1981) and Wollenzein and Cantor (1982). Noller et al. use the symbol "/" to mean "is base paired to". Because psoralen crosslinked nucleotides are not base paired to each other but rather to adjacent residues, we shall use the symbol "x" to denote "is crosslinked to". Wollenzein and Cantor have adopted a prefix to signify that their psoralen crosslinks

(Ps) have been localized by the electron microscope (E). The crosslinks described here have been localized using gel techniques (G). Thus, if bases 625 and 1420 are crosslinked by psoralen, this interaction is written as GPs 625 x 1420. In the following discussion, the three most widely cited secondary structure models will be called, for simplicity, the American model (Noller and Woese, 1981), the German model (Zwieb et al., 1981), and the French model (Stiegler et al., 1981).

Method of Crosslink Assignment

Assignment of crosslinks to specific bases in the following section has been done with a number of criteria in mind. First, and most obvious, is that the crosslink must occur within the isolated fragments. Possible secondary structures btween the two fragments were then determined by hand. These possibilities were discriminated on the basis of phylogenetic conservation and the presence of suitable psoralen crosslinking sites. Crosslinking sites were assumed to be between uridines only. While psoralen can certainly react with cytidine and even purines to some extent (Chapter 2), reaction with polymers has shown uridine to be the preferred site. Hot spots for monoaddition and crosslinking (Youvan and Hearst, 1982; Thompson et al., 1981) have all been at uridine. To some extent, looking for uridine-uridine crosslinks and then using these to prove the specificity of psoralen is a circular argument. However, the great ease in finding such sites argues in favor of this bias.

Phylogenetic comparisons were made using the best alignments available in the literature. The secondary structure alignments of Zwieb et al. (1981) and Stiegler et al. (1981) were found to be the most useful while the primary structure alignment of Kuntzel and Kochel (1981) was also helpful. The assignments below are separated into three categories: confirmatory crosslinks which are present in all models, discriminatory crosslinks present in some models, and new crosslinks.

Confirmatory Crosslinks

GPs 594 x 645: This crosslink confirms the presence of 588-617/623-651. While the presence of the helical structure is definite, the resolution of the crosslink does not allow an unambiguous assignment because there are two other likely crosslinking sites in the secondary structure shown in Figure 4. Also possible are GPs 598 x 641 and GPs 603 x 636. The secondary structure drawn is present in all three models in the literature. A UV-induced crosslink was also found in this region (Zwieb and Brimacombe, 1981).

GPs 1351 x 1372: This crosslink could also occur as GPs 1348 x 1375. 1350-1356/1366-1372 is present in all three models while 1347-1349/1376-1378 is found only in the German model. This helical region has also been crosslinked by UV light (Zwieb and Brimacombe, 1981).

GPs 252 x 273: This is the only likely site of crosslinking within the fragments found. It is near the end of a helix terminated by two A-U pairs so psoralen can easily intercal-

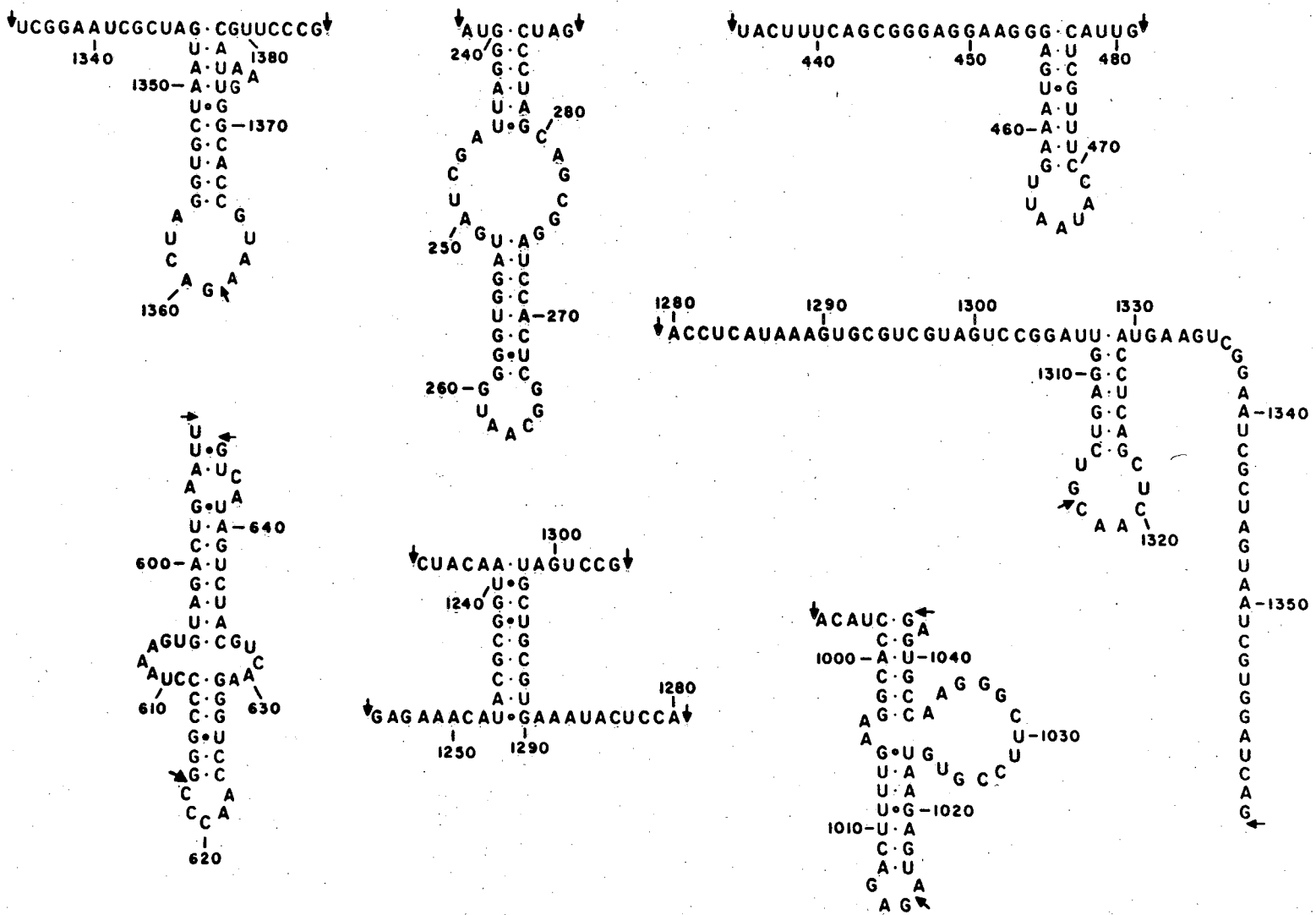

UCGGAAUCGCUAG·CGUUCCCG
1340
1380
1350
1360
1370
AUG·CUAG
240
250
260
270
280
UACUUUCAGCGGGAGGAAGGG·CAUUG
440
450
460
470
480
ACCUCAUAAAGUGCGUCGUAGUCCGGAUU·AUGAAGUC
1280
1290
1300
1310
1320
1330
1340
1350
CUACAA·UAGUCCG
1240
1250
1290
1300
GAGAAACAU·GAAAUACUCCA
1280
ACAUC·G
1000
1010
1020
1030
1040
620
600
610
630
640

Figure 4: Sequences of confirmatory crosslinks arranged in proposed secondary structures. The entire fragments of crosslinks present in published structure models are shown. Arrows mark the locations of observed T_1 cuts.

ate. The secondary structure proposed in all three models is identical in this region except for an additional G-U pair in the American model. The crosslink is isolated as a single fragment despite the presence of accessible guanines in the hairpin loop.

GPs 458 x 473: This fragment was also isolated as a single fragment. The three consecutive uridines opposite a G-U pair are ideal for crosslinking. This helix is well conserved in all species which have not deleted the region and has been shown to be near the bound mRNA (Wagner et al., 1976). This same psoralen crosslink has also been mapped by Turner et al. (1982) by a slightly different gel technique.

GPs 1007 x 1023: This crosslinking site is ideal for intercalation of psoralen. It contains a run of four uridines with a G-U pair at the end of the helix. This helix is present in all models and is supported by compensting base changes in other species.

GPs 1240 x 1298: This crosslinking site is at the base of an extended helix present in all three models. There are two potential psoralen sites within the proposed helix which are virtually identical and thus cannot be distinguished.

GPs 1308 x 1330: The hairpin stem proven by this crosslink is present in all models. The only uridine-uridine crosslink possible involves the terminal uridine on one strand with the first uridine beyond the helix on the opposite strand. This type of interaction might even be preferred over normal intercalation. There would be no unwinding of

the helix necessary and nearly as much stabilization by stacking would be gained. Stacking from the terminal base pair would be normal and some would also be gained from the next unpaired residues. Nucleotide 1308 was previously observed to be a hot spot for monoaddition by Youvan and Hearst (1982).

Discriminatory Crosslinks

GPs 1116 x 1183: The German model contains no sites of potential crosslinking that would generate the observed fragments while the newest version of the American model (H. Noller, personal communication) and the French model have two each. The most likely site in the American model (Figure 5) involves the terminal uridine in a helix with a uridine in a G-U pair. Also possible is a crosslink between terminal uridines in a coaxial stack (GPs 1118 x 1183). This helical region is deleted in some mitochondria but conserved in other species. A very similar base pairing scheme is present in the French model. In this model, GPs 1115 x 1183 would involve the terminal uridine in a helix with the first uridine beyond the helix. This type of crosslink has been proven for GPs 956 x 1506 and is likely for GPs 1308 x 1330. The other possible crosslink would arise if the helices formed by 1118-1124/1149-1155 and 1063-1067/1184-1193 stacked on each other. The terminal uridines would then be suitably positioned for crosslinking. These four crosslinks cannot be distinguished on the basis of the crosslinking data available.

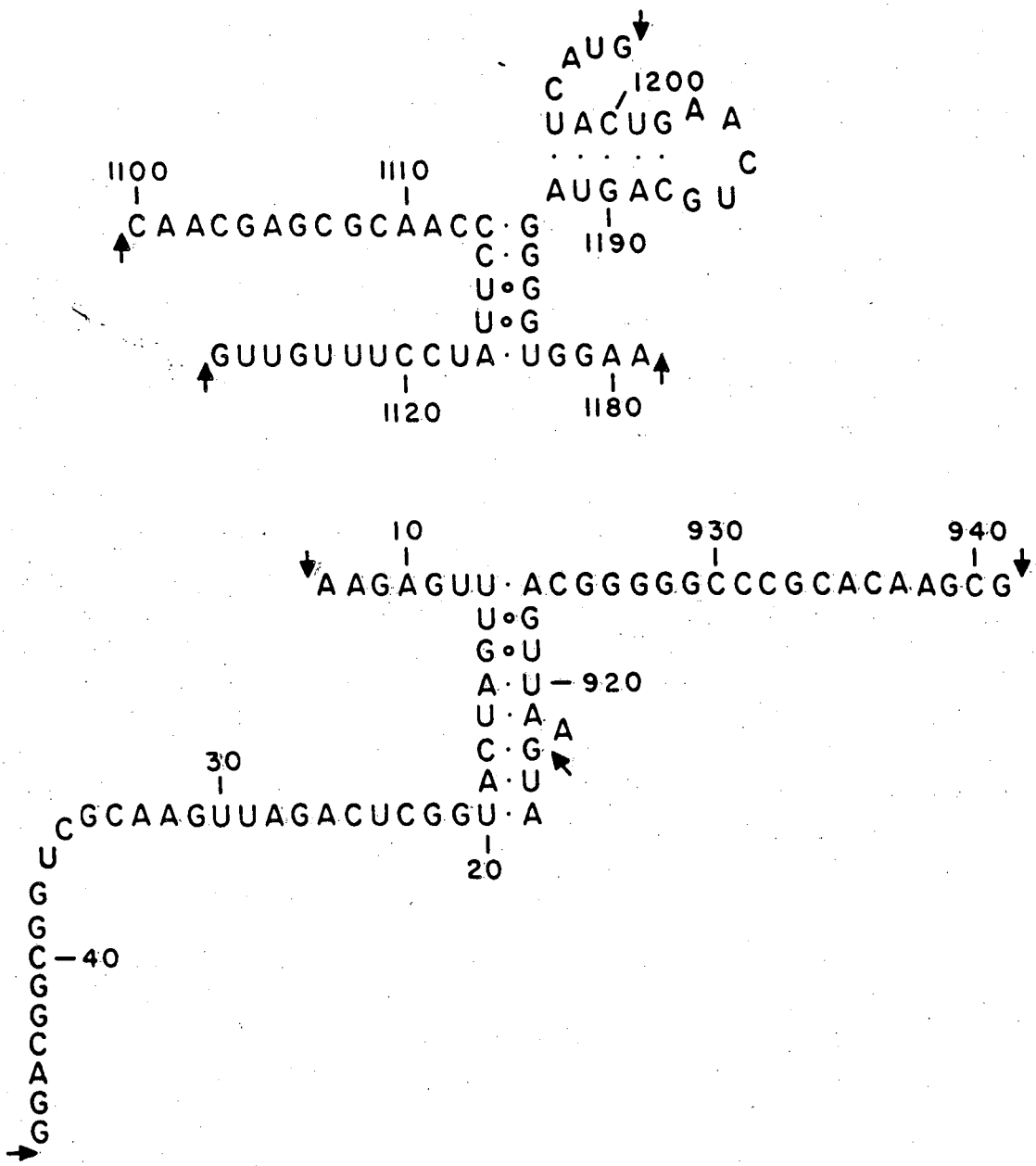

AUG
C 1200
UACUG A A
. C
AUGAC G U
1190
1100
1110
CAACGAGCGCAACC·G
C·G
U∘G
U∘G
GUUGUUUCCUA·UGGAA
1120
1180
10
930
940
AAGAGUU·ACGGGGGCCCGCACAAGCG
U∘G
G∘U
A·U
U·A
C·G
A·U
920
30
CGCAAGUUAGACUCGGU·A
20
C
U
G
G
C
G
G
C
A
G
G
40

Figure 5: Sequences and secondary structures of discriminatory crosslinks. Arrows show the positions of T_1 cuts. Three nucleotides have been added to the observed fragments of GPs14 x 921 in order to show the entire helical stem.

GPs 14 x 921: This crosslink must be classified as tentative because of problems in identifying one strand. The sequence for 918-941 is definite, however, 7-46 could not be read clearly enough to be sure of its uniqueness. Since nucleotides 920 and 921 are the only uridines in the first sequence, one of them is most likely involved in the crosslink. If 17-20/ 915-918, present in the American model, is extended by four base pairs with a single base bulge, a good crosslinking site would be present. This crosslink may have been observed in early electron microscopic work (Wollenzein et al., 1979) and erroneously classified with EPs 530 x 1540 The latter crosslink has been confirmed with a more rigorous polarity assignment recently (Wollenzein and Cantor, 1982) so the presence of this crosslink does not contradict the original work.

GPs 1189 x 1202: The helical region identified by this crosslink was present in two stretches in the original American model but has since been deleted (H. Noller, personal communication). It was not present at all in the French or German models. If GPs 1189 x 1202 is to occur simultaneously with 1116 x 1183, the helices in the original American model need to be shortened as shown in Figure 5.

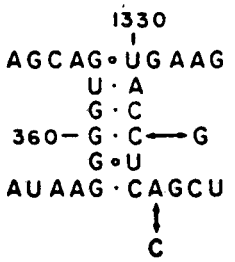
New Crosslinks

GPs 358 x 1330: While there are two stretches of complimentarity between the crosslinked fragments found (330-340/1333 1343 and 357-362/1325-1330), only one of these is conserved in other species as well (Figure 6). In this region, there

358 – 1330

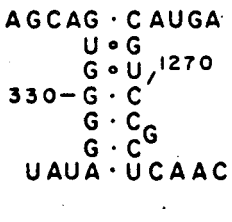
Prokaryotes

E. coli



(−9.8 kcal/mole)

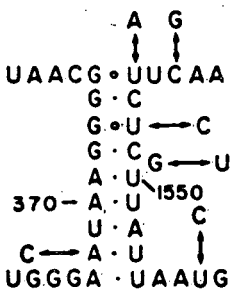
Z. mays (chloroplast)



(−11.6 kcal/mole)

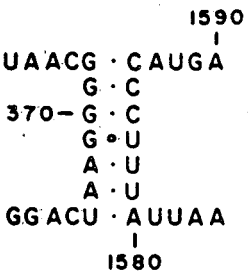
Eukaryotes

S. cerevisiae



(−9.0 kcal/mole)

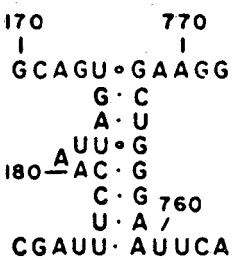
X. laevis



(−14.2 kcal/mole)

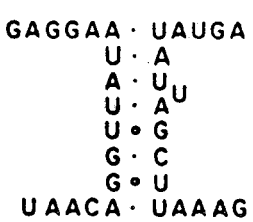
Mitochondria

Human



(−5.5 kcal/mole)

Yeast



(−5.6 kcal/mole)

Figure 6: Phylogenetic conservation of GPs 358 x 1330. The homologous secondary structures for 6 species of small subunit RNAs are shown. Base changes in going from the species on the left to the one on the right are shown by arrows.

is only one likely crosslinking site. In the American and German models, both of these regions are involved in other interactions. In the French model, one of the regions is involved in another interaction. This crosslink had been observed previously in the electron microscope and mapped as EPs 353 x 1344 (Cantor et al., 1980).

In order to conserve this interaction in Z. mays chloroplasts, one strand must be offset by two bases and a one base bulge introduced. One additional base pair can be made and more G-C pairs are present so this structure is actually more stable than in E. coli. Similar structures can be drawn for the homologous regions in eukaryotic RNAs. In mitochondria, however, only very weak interactions are found. Human and yeast mitochondria are only stable by 5.5 and 5.6 kcal/mole while no reasonable structure at all is present in mouse mitochondria. Secondary structure calculations were done by the method of Tinoco et al. (1973) using the most recent values available (I. Tinoco, Jr., personal communication).

It does not seem likely that coaxial stacking would generate this interaction. 368-379/384-393 has a suitably placed terminal uridine but there is no structure between 1317 and 1343 with a terminal uridine except for 1301-1305/1335-1339. This is present only in the German model and would necessitate stacking within a bulged loop of an extended helix.

GPs 625 x 1420: Overlapping T_1 fragments have allowed the

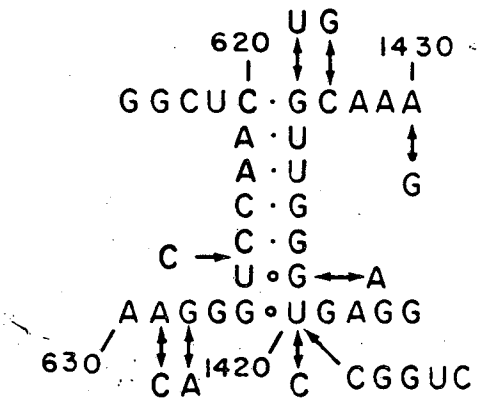
mapping of this crosslink to high resolution. There is only one good crosslinking site in the secondary structure shown (Figure 7). A similar structure is present in Z. mays chloroplasts. One base is bulged but additional G-C pairs more than make up for this. Deciding which structures are homologous is difficult in eukaryotes. Most of the additional sequences which have been inserted in 18S RNA are added between bases 600 and 640 of E. coli. Because of this, no region is strictly homologous to bases 620-626. There is no such difficulty with the other crosslinked fragment. When the inserted sequence is scanned for complementarity with the known homologous region, one stretch can be found that has good base pairing. This is shown in Figure 7. Finding the appropriate secondary structures for mitochondrial RNAs is much simpler. Homologous regions for both crosslinked strands have been deleted. This crosslink may correspond to Feature X found by electron microscopy and mapped at EPs 740 x 1370 (Wollenzein et al., 1979).

GPs 956 x 1506: This crosslink is the most accurately known of those described. Not only were the fragments in Table I found; but, using a technique involving complete T_1 digestion followed by running in a 2D gel system similar to that described by Turner et al. (1982), the exact T_1 fragments were found. This interaction is not present in any of the current models and would, in fact, necessitate the melting of one of the most highly conserved features of small subunit RNAs, the $m_2^6Am_2^6A$ hairpin structure. The very high de-

625-1420

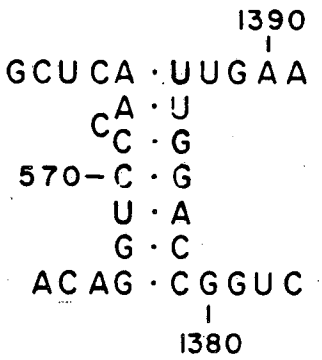
Prokaryotes

E. coli



(−11.8 kcal/mole)

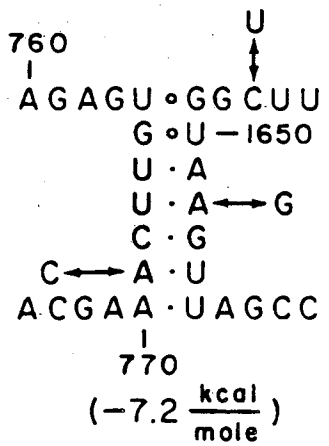
Z. mays (chloroplast)



(−14.1 kcal/mole)

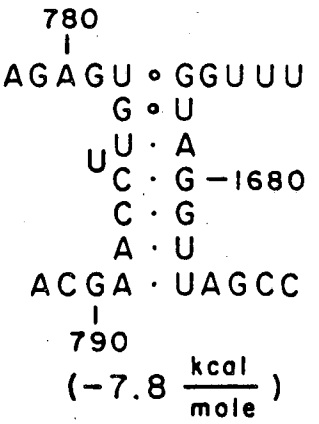
Eukaryotes

S. cerevisiae



(−7.2 kcal/mole)

X. laevis



(−7.8 kcal/mole)

Mitochondria

Human

Yeast

Figure 7: Phylogenetic conservation of GPs 625 x 1420.

gree of conservation of the interaction is shown in Figure 8 indicates that this helix is truly present in vivo and is not an artifact caused by the crosslinking or the conditions used. The significance of this structure will be discussed later. This feature was mapped by electron microscopy as EPs 970 x 1530 and was the most prevalent crosslink found (Wollenzein et al., 1979).

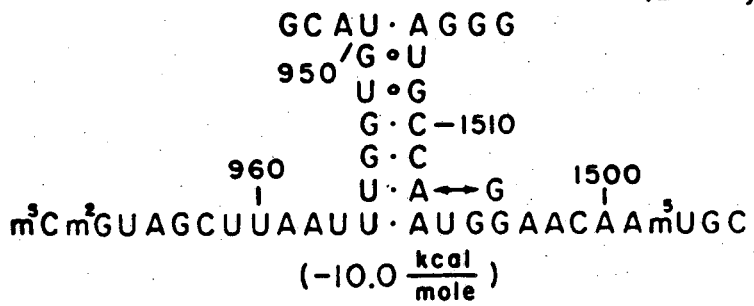
Dynamics of Long Range Crosslinks

The most surprising aspect of the long range interactions which have been observed is the presence of multiple structures for the same nucleotides. These structures are probably in equilibrium in the conditions used. In the functioning ribosome, such conformational switches are probably used to produce the physical movement necessary in translation. Evidence for switches has been found by Glotz et al. (1981). Psoralen may react particularly well with this type of structural feature because, by necessity, the helices must be relatively free to move and unwind. Since the helices are not tightly constrained by other parts of the RNA, psoralen is able to intercalate and crosslink.

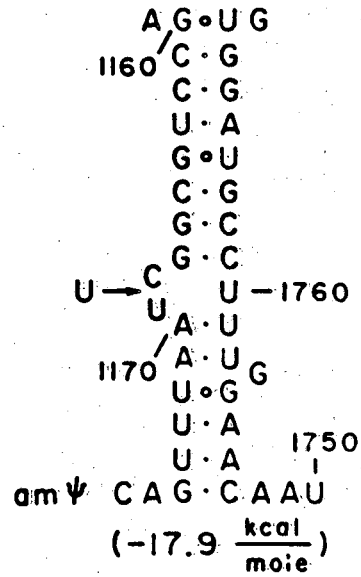
Two of the crosslinks found, GPs 1189 x 1202 and GPs 1116 x 1183 are in a region which there is evidence that large conformational changes occur (Glotz et al., 1981). The data from crosslinking and denaturation studies could not all be incorporated into a single secondary model. When our data is combined with that of Glotz et al.,(1981), it is quite apparent that large conformational changes are requir-

956-1506

Prokaryotes — E. coli (Z. mays)



Eukaryotes — S. cerevisiae (X. laevis)



Mitochondria — Human (Mouse)

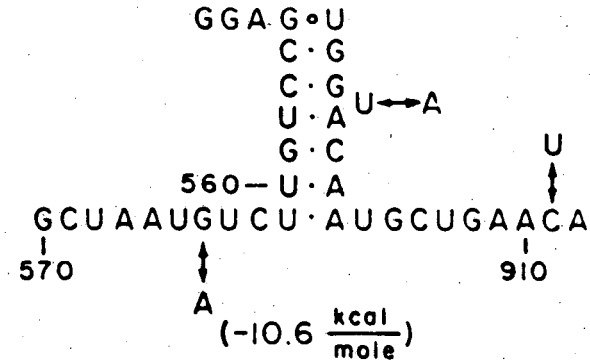
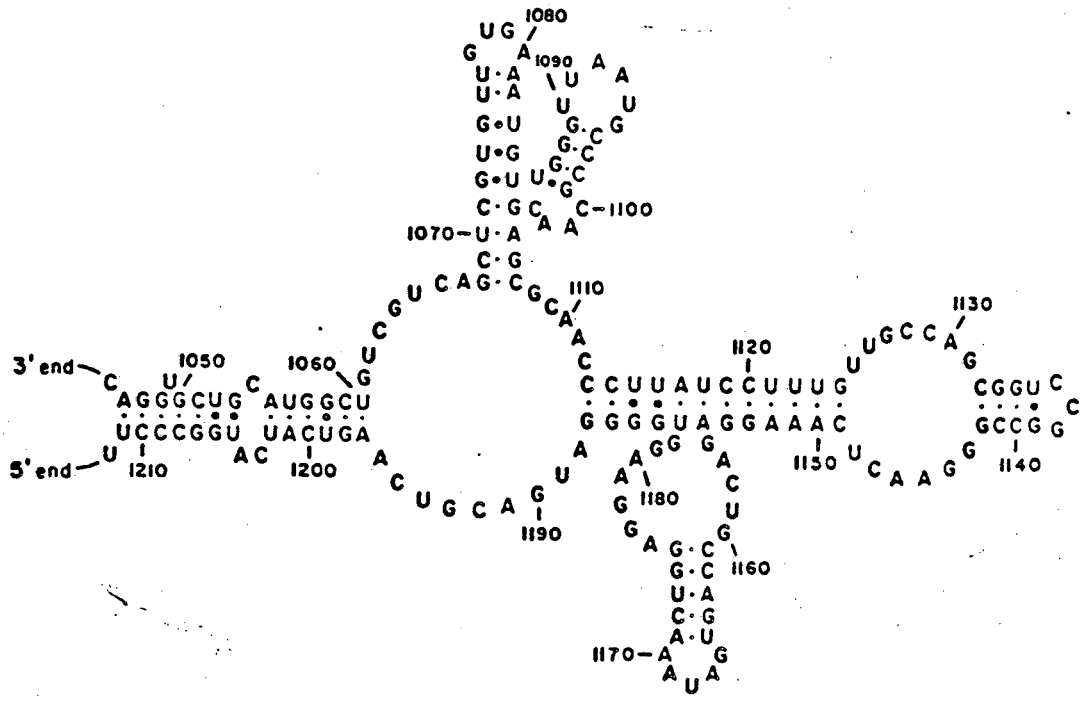


Figure 8: Phylogenetic conservation of GPs 956 x 1506.

ed to explain it. One possible set of structures is shown in Figure 9.

Possibly linked to this structural change is the crosslink GPs 358 x 1330. This crosslink is between regions similar to those in one interaction postulated by Glotz et al. (1981). The two interactions could easily co-exist and provide a stable link between the two regions of the RNA. Extensiveconformational changes can be drawn which combine the data of all three models as well as the evidence presented here. Whether the changes are as extensive as depicted in Figure 9 are very much in question. If the helices are as extended as pictured, topological problems could arise because more than a full helical turn would be introduced. GPs 1308 x 1330 must also be in dynamic equilibrium because it obviously could not co-exist with GPs 358 x 1330.

GPs 625 x 1420 also appears to be in a helix which undergoes conformational change. The formation of the helix shown in Figure 7 would require the opening of two other helices. The conformational change may be facilitated by the binding of protein S8. S8 has been shown to strongly protect 588-603/636-651 when bound (Ungewickall et al., 1975) and also induces a large conformational change when bound (Nomura et al., 1970). This change could very well be the bringing together of the regions around 620 and 1420. S8 does not necessarily have to bind directly to the region around 1420; instead, this interaction could be mediated by another protein.


3'end
5'end
1050
1060
1080
1090
1100
1110
1120
1130
1140
1150
1160
1170
1180
1190
1200
1210

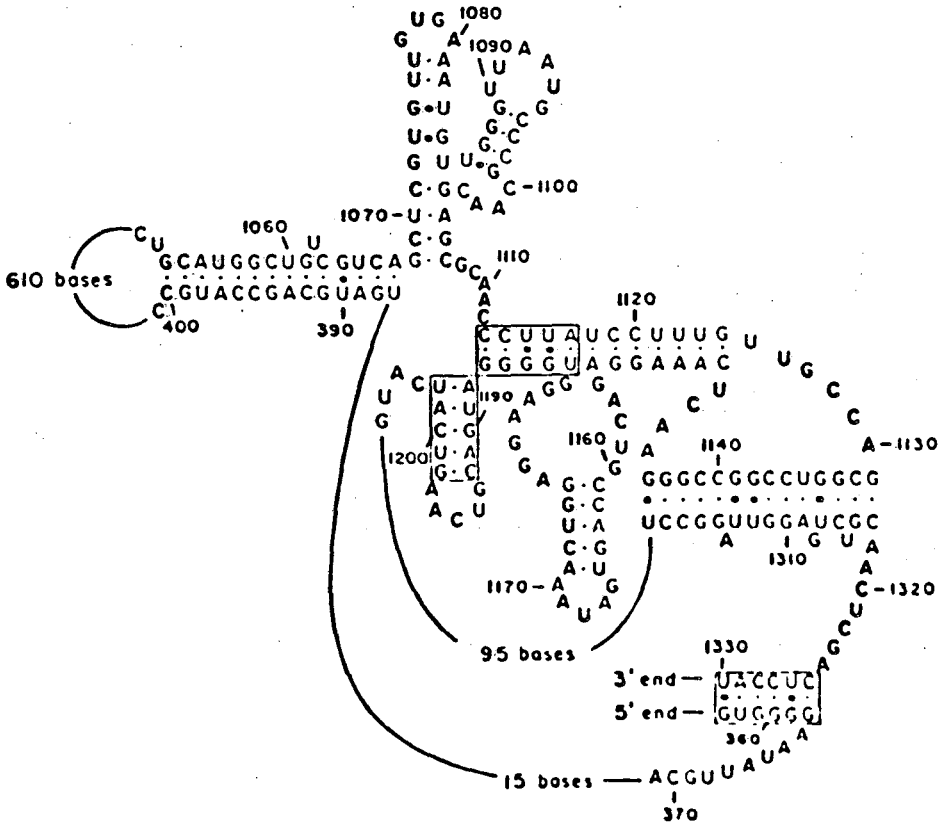

610 bases
95 bases
15 bases
3'end
5'end
1060
1070
1080
1090
1100
1110
1120
1130
1140
1160
1170
1190
1200
1310
1320
1330
1340
1350
1360
370
390
400

Figure 9: Possible conformational changes in 16S RNA. A) Long range interactions shown are supported by three different psoralen crosslinks (boxed) and one interaction found by Glotz et al. (1981), 387-400/1053-1067. 1131-1144/1301-1317 is not present in any of the current models but similar structures can be drawn for other species. B) Short range interactions shown are from the newest version of the American model (H. Noller, personal communication).

Additional evidence for this conformational switch can be deduced from the work of Stark et al. (1982). They have studied the effect of deletion mutations on the processing and function of E. coli 16S RNA. One of these mutations, a single base deletion at position 615, has a profound effect. Very little of this RNA is processed correctly or incorporated into 30S subunits. While this base is not directly involved in the long range interaction described, it would have an effect on the equilibrium of 612-617/623-628 $\rightleftharpoons$ 620-626/1420-1426. The stability of the first helix and hairpin loop is 12.0 kcal/mole while the second helix is 11.8 kcal/mole. Omitting base 615 does not change the stability of the second helix but the first structure is then only stable by 10.4 kcal/mole. Instead of being more stable, it is considerably weaker than the second helix after deletion. The short range interaction may be required for correct processing while the long range interaction would form with only a small input of energy at some other point in the ribosomal cycle. The same type of finely tuned equilibrium can be seen in Z. mays chloroplasts. The homologous helices have stabilities of 14.6 and 14.1 kcal/mole. Eukaryotes behave similarly with S. cerevisiae having stabilities of 7.4 and 7.2 kcal/mole and X. laevis having stabilities of 8.7 and 7.8 kcal/mole. While the sequence and stabilites of these helices in the four species vary widely, the trend of the short range interaction being slightly stronger than the long range interaction is uniform.

In the above analysis, the German secondary structure model has been used in all cases. The stabilities were calculated for the entire helix, including the hairpin loop, from where the disruption caused by the long range interaction would occur. Additional stability caused by repiring of the other two sequences involved in the short range interactions has not been included because there is no information on this. Disruption of the 1409-1430/1470-1491 helix has not been calculated either. This contribution is small because, of the seven base pairs needed to be broken for E. coli, four are G-U pairs and two are A-U pairs. Whether this structure is present as drawn in the German model is also subject to debate.

The helix crosslinked by GPs 956 x 1506 would also need to be in dynamic equilibrium. It would require the unpairing of the highly conserved $m_2^6Am_2^6A$ hairpin loop, a structure found in all small subunit RNAs that have been even partially sequenced. That this helix should open is not surprising in view of the results of Van Charldorp et al. (1981). They found that the presence of the four methyl groups destabilized the helix because of the greater energy required to unstack them for placement in a hairpin loop compared to unmodified adenines. This destabilization would not be present in 950-956/1507-1513 because this helix would not introduce a loop hence the two m_2^6As could effectively stack.

Functional Implications of Crosslinks

mRNA Binding: The role of 16S RNA in recognizing and bind-

ing mRNA in the initiation complex has been well established for several years (Shine and Dalgarno, 1975; Steitz and Jakes, 1975). In eukaryotes, however, the sequence which has been implicated in mRNA binding has been deleted. Because both eukaryotic and prokaryotic small subunits perform essentially the same function, there should be some compensating interaction between 18S RNA and mRNA. There are no apparent similarities in either the primary or secondary structure of eukaryotic mRNAs that would provide a basis for this.

GPs 956 x 1506 brings together two highly conserved regions in E. coli 16S RNA. The same interaction occurs in eukaryotes and thus brings this region spatially close to where mRNA recognition occurs in prokaryotes. In prokaryotes and eukaryotes, there are a number of modified bases located in both these parts of the RNA. As mentioned earlier, the two m_2^6As serve to facilitate the opening of the hairpin loop. In E. coli, there are a m^2G and a m^5C present in the region 550 bases from the 3' end. In eukaryotes, these have been replaced by the hypermodified base am$^{\psi}$. This nucleotide is perfectly suited to assist in mRNA recognition.

The modifications present on am$^{\psi}$ allow it to make specific interactions with the m^7G cap structure found at the 5' end of all eukaryotic mRNAs. The negative charge delocalized on the carboxylic acid group of am$^{\psi}$ can stabilize the positive charge delocalized on the imidizole nitrogens.

Simultaneously, the amino group of am^{Ψ} can interact with one of the negatively charged phosphate groups. Preliminary evidence suggests that exogenously added am^{Ψ} can inhibit translation (Chapter 7) as efficiently as exogenously added m^7GTP (Hickey et al., 1977) which lends additional support to the importance of this interaction. A more detailed description of this interaction and additional evidence for it are presented in Chapter 7.

This interaction may also be involved in more complex intersubunit contacts. Azad (1979) has proposed an interaction between 5S RNA and the same region of 16S RNA (1509-1517) that pairs with the region near m^2Gm^5C. There is no firm evidence for this interaction and it was suggested by Schnare and Gray (1981) that it is not universal. However, stable base pairing of 5S and 18S RNA in solution has been observed (Oakden et al., 1976). The in vitro complex formed between D. melanogaster 18S and 5S RNAs can be crosslinked by HMT and large amounts of 5S co-purify with D. melanogaster 18S RNA even after two rounds of sucrose gradient centrifugation using the purification protocol described in the Materials and Methods section. The fact that 30S subunits which contain the 956-1506 crosslink are less able to form 70S ribosomes than other crosslinked subunits (Thammana et al., 1979) further suggests that 5S pairs with 16S through this interaction.

How the two long range and the short range interactions might alternate through the ribosomal cycle is not clear.

For instance, one interaction might only occur during initiation while the others might switch during elongation. Only crosslinking results from ribosomes irradiated at specific points in translation will clarify this situation.

Proofreading and tRNA Binding: The total error rate in translation is simply a sum of the error rates of its component reactions. The theoretical and practical problems involved in the analysis of translational fidelity are reviewed by Kurland (1980) and Yarus (1979). Best estimates place the total error rate from all factors at one misincorporation per 10^4 amino acids. The only step in translation which cannot be expected to easily yield this level of discrimination is tRNA binding via the codon-anticodon interaction. The difference in binding energies of partially degenerate anticodons is far too small to expect such accurate reading. To account for this, a number of models have been presented, all of which involve reading the anticodon twice to multiply small differences in binding. The lack of experimental data has, up until now, prevented formulation of a detailed physical model of this process which satisfactorily accounts for what little is known.

No part of 16S RNA has been associated with a proofreading function. Several proteins, however, are known to be involved in regulating translational fidelity. Elongation factor Tu, S4, S5, S11, S12, and S17 have all been shown to profounly affect the error rate (Gavrilova et al., 1981 and references therin). The characteristics of one of

the long range crosslinks observed, GPs 625 x 1420, suggests that it might have a role in proofreading and tRNA binding. The region near 1420 has been implicated in binding of tRNA to the P site (Taylor et al., 1981) and the region near 625 is part of the S8 (a tRNA binding protein) binding site. Both regions are highly variable as would be expected for a proofreading function. The proofreading process requires energy and each species will have different requirements for optimizing the advantages of increased accuracy with the disadvantages of energy loss. Thus, proofreading should be different even among closely related species with even larger differences upon going from mitochondria to prokaryotes to eukaryotes. The region of E. coli 16S RNA around 580-660 varies considerably among prokaryotes and has been deleted entirely by mitochondria. Virtually all of the nucleotides which have been inserted into eukaryotic 18S RNA are found in this region. A similar behavior is seen around 1420. Mitochondria have shortened that helical stem while eukaryotes have expanded it.

Intuitively, one would expect eukaryotes to require the lowest error rate and hence devote more of the 18S RNA to that task. Eukaryotes synthesize many more proteins than prokaryotes and are thus more sensitive to error-induced damage. Mitochondria, on the other hand, are almost free of proofreading constraints. All proteins synthesized are multiple copy and only a few different ones are made. Indeed, all proteins which could propagate errors, ribosomal pro-

teins and polymerases, are synthesized outside the mitochondria. In some cases, mitochondria only read two of the anticodon nucleotides (Heckman et al., 1980) so it would not be surprising if they were to delete all or part of the proofreading apparatus. This analysis has not relied on experimental results because measures of in vivo translational fidelity are extremely difficult to do and have been restricted to studies which detect only one or a few different misincorporated amino acids in a protein (Edelmann & Gallant, 1977). Even these studies have to be viewed critically because E. coli ribosomes can reject nascent peptides which contain an error (Caplan & Menninger, 1979). These oligopeptides are broken down rapidly in the cell so are difficult to quantitatively measure.

Lake (1979) has proposed a detailed model for what he terms the R (recognition) site of tRNA binding. The anticodon is read once in the R site. A conformational change in the tRNA occurs to bring it to the A site where the anticodon is read again. A primary reason for placing the R site on the exterior of the 30S subunit is the location of several tRNA binding (including S8) and proofreading proteins there. For the reasons mentioned earlier, the cyclic interaction of 612-617/623-628⇌620-626/1420-1426 appears to be ideally suited for involvement in the process of moving a tRNA from the R site to the A site.

Lake (1981) proposes that the conformational change which brings the tRNA to the A site would occur solely in

the tRNA with the only tRNA contact to the ribosomal complex being at the anticodon. This seems unlikely not only because of the weakness of some codon-anticodon interactions, but also because of the ease with which the process could be short-circuited. If the tRNA in the process of switching were to come off the mRNA, there would be nothing to prevent a new tRNA which had not undergone the initial screening at the R site from taking its place and moving into the A site. It is more likely that there are multiple tRNA-protein and tRNA-rRNA contact points which ensure that the bound tRNA has all the important features of the cognate aminoacyl tRNA. In this way, other conformational changes in the ribosome could be tightly coupled to tRNA movement.

Ef-Tu, which has been shown to recognize the 3' end of aminoacyl tRNA before binding the ribosome (reviewed by Weissbach, 1980), and the tRNA binding proteins on the exterior of the 30S subunit would make contact with tRNA bound to the R site. It would not be surprising if rRNA were also involved. The high variability of the 588-617/623-651 region suggests that it would not be directly involved in tRNA binding. There are, however, two nearby sequences of CGAA that are highly conserved. Both of these stretches, located at 726-729 and 764-767 in E. coli, are present in all prokaryotes and eukaryotes and at least one is present in all mitochondria. CGAA is complementary to the highly conserved TΨCG present in tRNA. While this sequence is not available for inter-RNA binding in solution, there is strong evidence

that binding of a codon to tRNA makes this region more accessible (reviewed by Kim, 1978) and thus able to bind to 16S RNA or 5S RNA. Such an interaction would destroy contact between the D and TΨ loops of tRNA, also freeing other sequences for interaction. The conserved YGG sequence in the D loop could also be involved in interactions in the R site or it might remain free to allow specific binding to the A site upon switching.

There is a strongly conserved sequence in 16S RNA which would allow pairing of the YGG sequence in the A site and subsequently in the P site. The sequence CCGm^{4}CmCCG (1300-1405 in E. coli) is present in all prokaryotes and eukaryotes and some mitochondria. In the A site, 1399-1401 would pair with the exposed YGG while the P site tRNA YGG could pair with 1403-1405.

This hypothesis is based partly on the data of Ofengand et al. (1982). They found a modified base in the anticodon loop of a P site tRNA could crosslink to C-1400. This crosslinking was done with an empty A site so C-1400 would be available. When the proper codon was supplied for the tRNA (Ofengand and Liou, 1981), crosslinking was abolished. This indicates that the interaction between the anticodon and C-1400 is probably not functionally important. It does, however, establish that the P site is in very close proximity. The A site must also be very close because Johnson et al. (1982) found that the distance between the s^{4}U position in tRNAs bound to the A and P sites is only 2-10 Å greater

than the tRNA diameter. This implies that the A and P site tRNAs are in very close contact throughout their entire lengths because the anticodon loops and 3' ends also have to be quite close.

In order for the D loops to pair with 16S RNA as described, the tRNA would have to undergo a conformational change. The D and TΨ loops would have already separated in the R site. The D loop and stem would also have to twist slightly and move toward the anticodon loop. In such a conformation, the bases in the anticodon and D loops are in an antiparallel configuration. The A site is on the 3' side of the mRNA so it is on the 5' side of the 16S RNA. A diagram of how this might happen is shown in Figure 10. The mRNA is necessarily kinked so the anticodon loops can base pair to adjacent codons on the mRNA. The same type of kink is necessary in the 16S RNA. There is a larger distance allowable between the D loops because the base paired regions are separated by an unpaired, modified nucleotide. The base and sugar methylations might somehow stabilize this kink.

Except for a few mitochondrial tRNAs and tRNAs involved in cell wall synthesis, the sequence GG is present in the same location in the D loop. The base 5' to this is usually a C, D, or U but occasionally an A. The pairing of G-1401 or G-1405 with this variable base holds additional potential for distinguishing between different tRNAs. The structure of G-C, G-U, G-D, and G-A pairs are all different and may change the orientation of a tRNA enough that, depending on

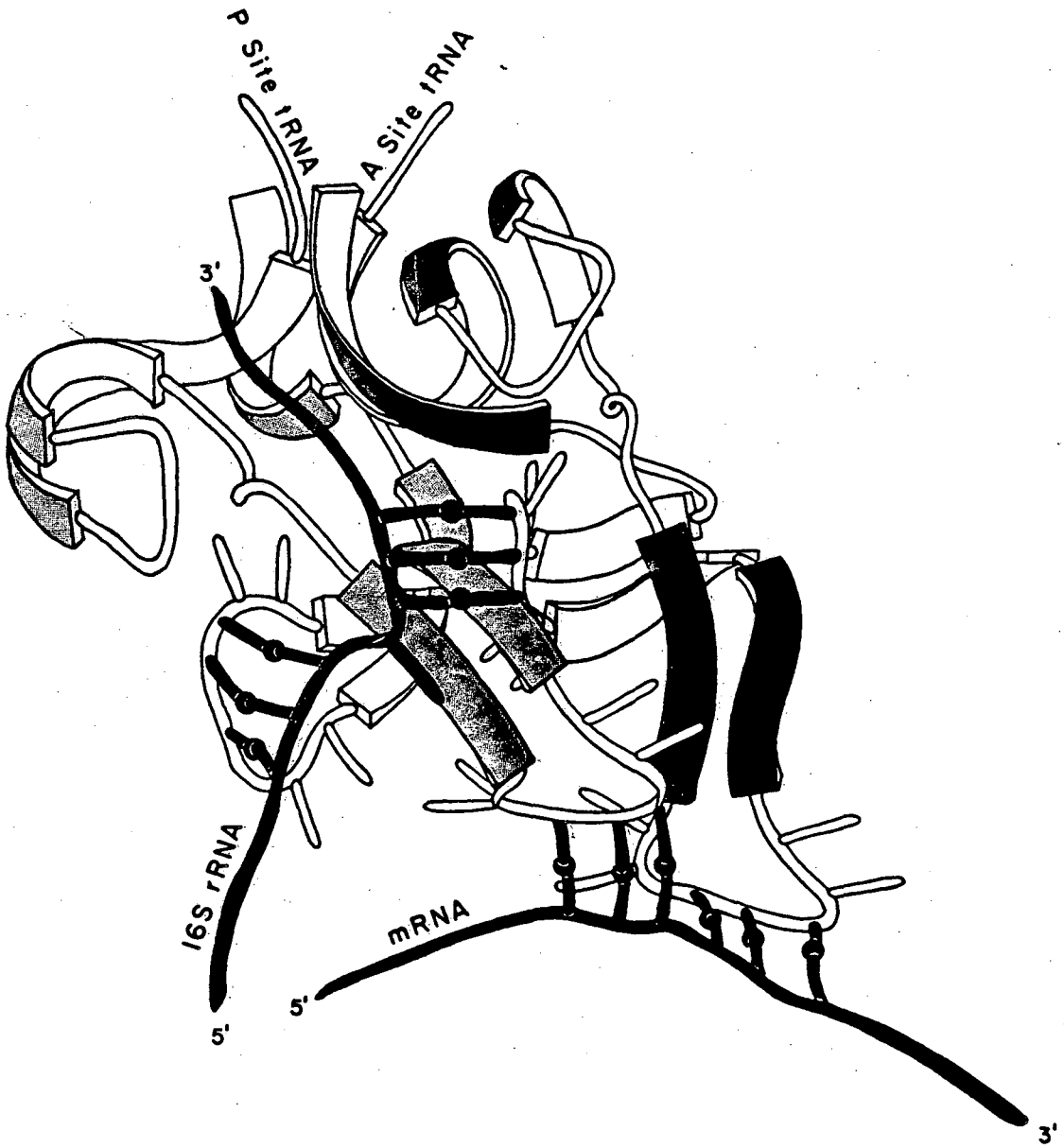

P Site tRNA
A Site tRNA
3'
16S rRNA
5'
mRNA
5'
3'

Figure 10: Possible structure of two tRNAs interacting with both mRNA and 16S RNA in the A and P sites of the ribosome.

the remainder of the structure, it could stabilize cognate and destabilize non-cognate tRNAs.

Translocation and Elongation: Very little is known about the mechanism of translocation and which parts of the ribosome are involved. The following model for translocation and movement of tRNAs through the ribosome is presented not as a definitive statement on how the ribosome works but as a way of accounting for our observations and those of other workers in the field. There is clearly much work to be done and this model should help point out weaknesses in our knowledge.

The role of tRNAs in translocation appears to be paramount. The distance which mRNA moves is determined by the tRNA (Thach and Thach, 1971; Gupta et al., 1971). Johnson et al. (1982) have also proposed that the energy for translocation comes from energy stored when the A site tRNA is tightly packed adjacent to the P site tRNA. In order for the tRNAs to be in such close contact, there must be other parts of the ribosome which prevent the tRNAs from escaping. Since this would necessarily have to be a cyclic interaction RNA-RNA interactions seem likely to be involved. While inter-subunit interactions could play a role in this, we have no information which relates to this. A large part of the tRNA binding sites are localized on the 30S subunits so 16S RNA could certainly play a major role.

Brimacombe (1980), while not setting forth a specific model, proposed that 39-47/393-402 + 1055-1065/1186-1195 →

385-399/1052-1067 might somehow be involved in translocation. There is no direct evidence for Brimacombe's proposal but this seems to be exactly the type of cyclic, long range interaction necessary if translocation is to proceed as described above. The interactions shown in Figure 9 are certainly intricate enough to lock the P site tRNA in place. At the opposite end of the 30S subunit, similar interactions would have to occur to lock the A site tRNA in place. These could include 950-956/1507-1513 and other interactions such as EPs 450 x 1540, EPs 510 x 1540, or EPs 0 x 1540 which have been mapped by electron microscopy (Wollenzein et al., 1979; Wollenzein and Cantor, 1982) but not known to high enough resolution to describe in detail.

A chart showing our proposed model for the elongation process is shown in Figure 11. The key features and abbreviations used are explained in the figure legend. The role of elongation factors and conformational changes are shown with the occupancy of each tRNA binding site after each event is listed.

When EF-G ·GTP binds to the pretranslocation ribosome, it destabilizes the long range interactions and causes the short range base pairing of Figure 9B to occur. This provides an escape route for the P site tRNA. This tRNA is rapidly expelled from the P site because of electrostatic repulsion from the A stie tRNA. The A site tRNA moves to the P site simultaneously because of the much greater affinity of peptidyl tRNAs for the P site. After tRNA movement,

A Model for Protein Elongation

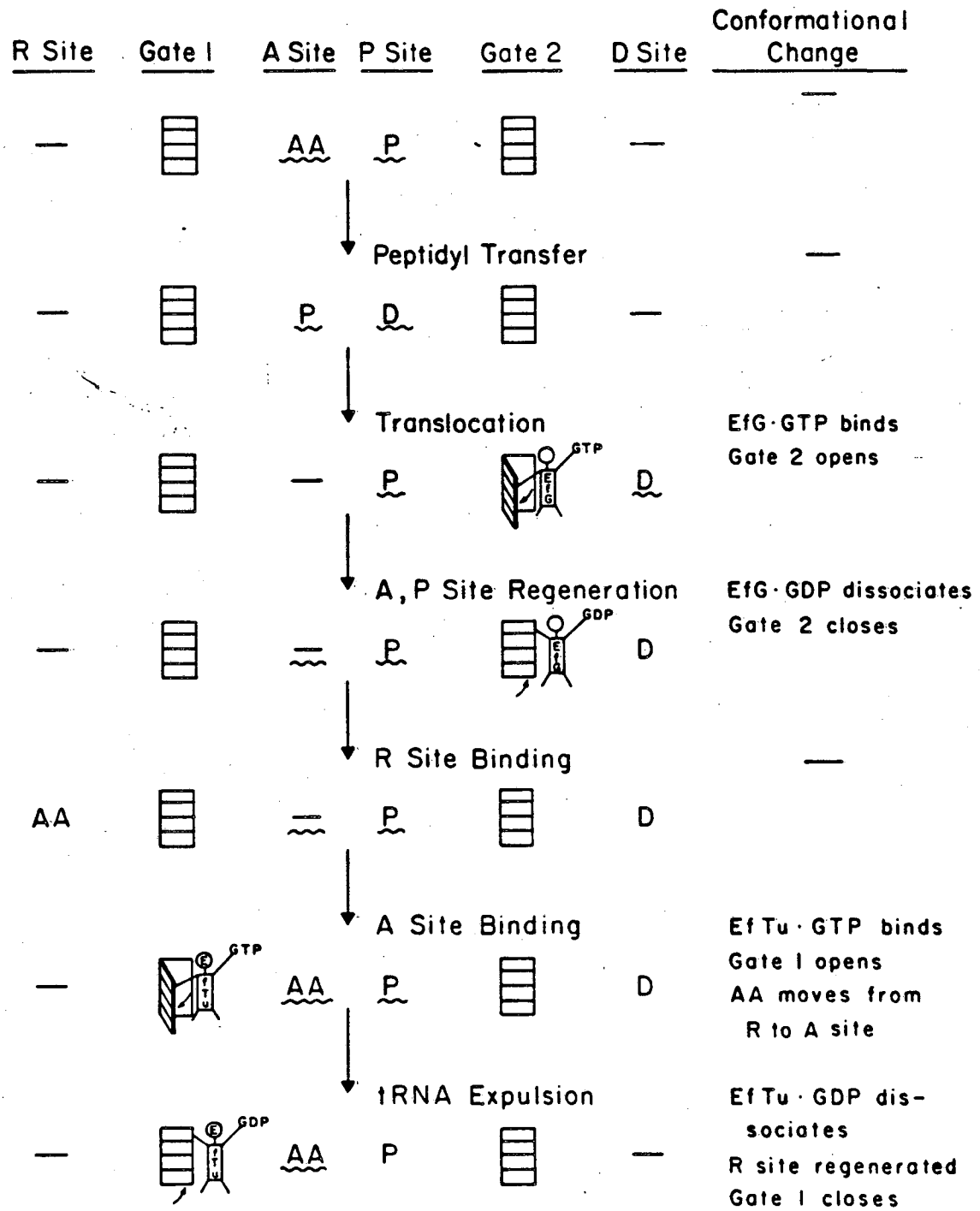


Figure 11: Model for elongation and translocation.

The model involves six steps, three of which involve structural transitions in the RNA. These transitions are described below.

1) Gate 1 is a steric barrier to tRNA movement at one end of the cleft in the 30S subunit. It separates the R and A sites and is associated with the following, and possibly other, transitions in secondary structure:

Gate 1 (closed) $\rightleftharpoons$ Gate 1 (open)

950-956/1507-1513 $\rightleftharpoons$ 946-955/1225-1235 + 1506-1515/1520-1529

2) Gate 2 is a similar barrier at the other end of the cleft between the P and D sites. It is associated with the transition described in Figure 9.

Gate 2 (closed) $\rightleftharpoons$ Gate 2 (open)

Figure 9A $\rightleftharpoons$ Figure 9B

3) The movement of the aminoacyl tRNA from the R site to the A site:

AA-tRNA (R site) $\rightarrow$ AA-tRNA (A site)

612-617/623-628 $\rightarrow$ 620-626/1420-1426

During the next step (tRNA expulsion), the reversal of this structural change in the 16S RNA occurs, leaving the aminoacyl tRNA in the A site and generating an empty R site.

In the above figure, four sites of tRNA binding are postulated. The R (recognition) site corresponds, in principle, to that described by Lake (1981). The physical attributes and position are not necessarily the same, however. The A and P sites are as usually proposed. The D (discharge)

site corresponds to the E site of Rheinberger et al (1981). The name has been changed for acronymic reasons. AA refers to the aminoacyl tRNA, P to the peptidyl tRNA, and D to the deacylated tRNA. ▮ refers to a closed gate, a blank space to an open gate, and ∿ to positions of the CCG sequences in the 16S RNA which base pair with the D loop of the tRNA (see Discussion).

EF-G·GDP dissociates from the ribosome, catalyzed by GTP hydrolysis. The stabilization by EF·G of short range interactions is no longer a factor as to which conformation is favored so the interdomain interactions shown in Figure 9A are re-established.

Evidence for a D site and tRNA binding to it after the P site is described by Rheinberger et al. (1981) (the E site has been renamed the D site for acronymic reasons). The presence of the deacylated tRNA in the D site accelerates the binding of the aminoacyl tRNA·EF-Tu·GTP ternary complex to the ribosome. This occurs after an initial reading of the anticodon of the incoming tRNA. Once the anticodon has been interpreted as correct, EF-Tu·GTP binds with high affinity to the short range interactions near the A site (946-955/1225-1235 + 1506-1515/1520-1529) and allows the R site tRNA to move into the A site. In the presence of EF-Tu, this movement is irreversible without added energy and provides the non-equilibrium situation necessary for true proof reading to occur (Yarus, 1979; Kurland, 1980). Once in place, the tRNA anticodon is reread. If still deemed correct, EF-Tu·GDP dissociates from the ribosome with hydrolysis of GTP. This allows the long range interactions to reform (950-956/1507-1513) and locks the two tRNAs in place, correctly positioned for peptidyl transfer.

This model is necessarily incomplete but does account for all the data available on elongation at present. For instance, while four tRNA binding sites are proposed, only 2

or 3 are occupied at any one time. This agrees with the data of Rheinberger et al. (1981) who found 2 to 2.5 tRNAs bound during translation. It also includes the R and D sites which increase the fidelity of translation (Lake, 1981; Rheinberger et al., 1981). The properties of non-cleavable GTP analogs in factor binding is accounted for because the energy input is used solely to drive dissociation. The factors simply favor one direction in a conformational equilibrium. Non-enzymatic translation is possible because the same conformational equilibrium would be present in the absence of factors but would simply occur at a slower rate.

Other Methods of Psoralen Crosslink Mapping: Electron microscopy has been the most powerful tool for localizing psoralen crosslinks up until now. This technique is limited primarily by resolution. Short range crosslinks cannot be observed at all while long range crosslinks cannot be mapped as accurately as described here. Now that the type of site that psoralen prefers to crosslink is known better, mapping can be done with somewhat more confidence. Other gel techniques have also been used (Thompson et al., 1981; Turner et al., 1982). While providing valuable information, these techniques are simply not as versatile as those described here.

Complete digestion of fragments with T_1 provides higher resolution but short fragments are sometimes impossible to place in the sequence. Complete T_1 digestion followed by reversal could be employed in finding GPs 956 x 1506 but

only because it was present in high yield (Wollenzein et al., 1979) and contained two relatively long T_1 fragments.

CONCLUSIONS

The large number of psoralen crosslinks which have been localized in this study clearly shows the usefulness of this technique for determining the secondary structure of any large RNA. The technique is also simple enough that most laboratories can readily employ it on any RNA of interest.

The results presented here, while supporting the generally accepted models for 16S RNA structure, also clearly point out the dynamic nature of the molecule. Only when we can correlate conformational changes with specific events will we begin to understand the ribosomal machinery. The value of psoralen crosslinking in future studies is high. Psoralen crosslinks precisely those features which are most difficult to observe using other techniques.

REFERENCES

Azad, A. A. (1979) Nucl. Acids Res. 7, 1913-1929.

Bachellerie, J. -P., Thompson, J. F., Wegnez, M. R., & Hearst, J. E. (1980) Nucl. Acids Res. 9, 2207-2222.

Bachellerie, J. -P. & Hearst, J. E. (1982) Biochemistry 20, 1357-1363.

Barnes, W. M. (1978) J. Mol. Biol. 119, 83-99.

Berger, S. L. & Birkenmeier, C. S. (1979) Biochemistry 18, 5143-5149.

Brimacombe, R. (1980) in Biological Implications of Protein-Nucleic Acid Interactions (ed. by J. Augustyniak), pp. 44-62, Elsevier-North Holland Biomedical Press, Amsterdam.

Brosius, J., Palmer, M. L., Kennedy, P. J., & Noller, H. F. (1978) Proc. Nat. Acad. Sci. USA 75, 4801-4805.

Caplan, A. B. & Menninger, J. R. (1979) J. Mol. Biol. 134, 621-637.

Cantor, C. R., Wollenzein, P. L., & Hearst, J. E. (1980) Nucl. Acids Res. 8, 1855-1872.

Carbon, P., Ehresmann, C., Ehresmann, B., & Ebel, J. -P. (1979) Eur. J. Biochem. 100, 399-410.

Carbon, P., Ebel, J. -P., & Ehresmann, C. (1981) Nucl. Acids Res. 9, 2325-2333.

Edelmann, P. & Gallant, J. (1977) Cell 10, 131-138.

Eperon, I. C., Anderson, S., and Nierlich, D. P. (1980) Nature 286, 460-464.

Fellner, P., Ehresmann, C., & Ebel, J. -P. (1970) Nature

225, 26-29.

Gavrilova, L. P., Perminova, I. N., & Spirin, A. S. (1981) J. Mol. Biol. 149, 69-78.

Glotz, C. & Brimacombe, R. (1980) Nucl. Acids Res. 8, 2377-2395.

Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K., & Kossel, H. (1981) Nucl. Acids Res. 9, 3287-3306.

Gupta, S. L., Waterson, J., Sopori, M. L., Weissman, S. M., & Lengyel, P. (1971) Biochemistry 10, 4410-4421.

Heckman, J. E. Sarnoff, J., Alzner-DeWeerd, B., Yin, S. & RajBhandary, U. L. (1980) Proc. Nat. Acad. Sci. USA 77, 3159-3163.

Hickey, E. D., Weber, L. A., Baglioni, C., Kim, C. H., & Sarma, R. H. (1977) J. Mol. Biol. 109, 173-183.

Hirsch, D. (1971) J. Mol. Biol. 58, 459-468.

Johnson, A. E., Adkins, H. J., Matthews, E. A., & Cantor, C. R. (1982) J. Mol. Biol. 156, 113-140.

Kanne, D., Straub, K., Rapoport, H., & Hearst, J. E. (1982) Biochemistry 21, 861-871.

Kim, S. -H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A., Seeman, N. C., & Rich, A. (1974) Proc. Nat. Acad. Sci. USA 71, 4970-4979.

Kim, S. -H. (1978) in Transfer RNA (ed. by S. Altman), pp. 248-293, M. I. T. Press, Cambridge, Mass.

Kuntzel, H. & Kochel, H. G. (1981) Nature 293, 751-755.

Kurland, C. G. (1980) in Ribosomes (ed. by G. Chambliss et al.), pp. 597-614, University Park Press, Baltimore.

Lake, J. A. (1979) in Transfer RNA: Structure, Properties, and Recognition (ed. by P. R. Schimmel et al.), pp. 393-411, Cold Spring Harbor Laboratory.

Lake, J. A. (1981) Sci. Amer. 245, 84-97.

Muller, A., Manderscheid, U., Lipecky, R., Bertram, S., Schmitt, M., & Gassen, H. G. (1979) in Transfer RNA: Structure, Properties, and Recognition (ed. by P. R. Schimmel et al.), pp. 459-471, Cold Spring Harbor Laboratory.

Nierhaus, K. H., Rheinberger, H. J., Wurmbach, P., & Bergemann, K. (1980) in Biological Implications of Protein-Nucleic Acid Interactions (ed. by J. Augustyniak), pp. 121-131, Elsevier-North Holland Biomedical Press, Amsterdam.

Noller, H. F. & Woese, C. R. (1981) Science 212, 403-411.

Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., & Herr, W. (1981) Nucl. Acids Res. 9, 6167-6189.

Nomura, M., Mizushima, S. Ozaki, M., Traub, P., & Lowry, C. V. (1969) Cold Spring Harbor Symp. Quant. Biol. 34, 49-61.

Oakden, K. M., Azad, A. A., & Lane, B. G. (1977) Can. J. Biochem. 55, 99-109.

Ofengand, J. & Liou, R. (1981) Biochemistry 20, 552-559.

Ofengand, J., Gornicki, P., Chakraburtty, K., & Nurse, K. (1982) Proc. Nat. Acad. Sci. USA 79, 2817-2821.

Rabin, D. & Crothers, D. M. (1979) Nucl. Acids Res. 7, 689-

703.

Rheinberger, H. -J., Sternbach, H., & Nierhaus, K. H. (1981) Proc. Nat. Acad. Sci. USA 78, 5310-5314.

Ross, A. & Brimacombe, R. (1979) Nature 281, 271-276.

Rubtsov, P. M., Musakhanov, M. M., Zakharyev, V. M., Krayev, A. S., Skyrabin, K. F., & Bayev, A. A. (1980) Nucl. Acids Res. 8, 5779-5795.

Salim, M. & Maden, B. E. H. (1981) Nature 291, 205-208.

Schwarz, Z. & Kossel, H. (1980) Nature 283, 739-742.

Schnare, M. N. & Gray, M. W. (1981) FEBS Let. 128, 298-304.

Sedat, J. W., Kelly, R. B., & Sinsheimer, R. L. (1967) J. Mol. Biol. 26, 537-540.

Shine, J. & Dalgarno, L. (1975) Nature 254, 34-38.

Sor, F. & Fukuhara, H. (1980) C. R. Acad. Sci. (Paris) Ser. D. 291, 933-936.

Stark, M. J. R., Ghourse, R. L., & Dahlberg, A. E. (1982) J. Mol. Biol., in press.

Stiegler, P., Carbon, P., Ebel, J. -P., & Ehresmann, C. (1981) Eur. J. Biochem. 120, 487-495.

Steitz, J. A., & Jakes, K. (1975) Proc. Nat. Acad. Sci. USA 72, 4734-4738.

Straub, K., Kanne, D., Hearst, J. E., & Rapoport, H. (1981) J. Am. Chem. Soc. 103, 2347-2355.

Taylor, B. H., Prince, J. B., Ofengand, J. & Zimmerman, R. A. (1981) Biochemistry 20, 7581-7588.

Thach, S. S. & Thach, R. E. (1971) Proc. Nat. Acad. Sci. USA 68, 1791-1795.

Thompson, J. F., Wegnez, M. R., & Hearst, J. E. (1981) J. Mol. Biol. 147, 417-436.

Thompson, J. F., Bachellerie, J. -P., Hall, K., & Hearst, J. E. (1982) Biochemistry 20, 1363-1368.

Tinoco, I. Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., & Gralla, J. (1973) Nature New Biol. 246, 40-41.

Traub, P., Mizushima, S., Lowery, C. V., & Nomura, M. (1971) Meth. Enzymol. 20, 391-401.

Turner, S., Thompson, J. F., Hearst, J. E., and Noller, H. F. (1982) Nucl. Acids Res. 10, 2839-2849.

Ungewickall, E., Garrett, R., Ehresmann, C., Stiegler, P. & Fellner, P. (1975) Eur. J. Biochem. 51, 165-180.

Van Etten, R. A., Walberg, M. W., & Clayton, D. A. (1980) Cell 22, 157-170.

Wagner, R., Gassen, H. G., Ehresmann, C., Stiegler, P., & Ebel, J. -P. (1976) FEBS Let. 67, 312-315.

Weissbach, H. (1980) in Ribosomes (ed. by C. Chambliss et al.), pp. 377-411, University Park Press, Baltimore.

Wollenzein, P. L., Hearst, J. E., Thammana, P., & Cantor, C. R. (1979) J. Mol. Biol. 135, 255-269.

Wollenzein, P. L. & Cantor, C. R. (1982) J. Mol. Biol. 159, 151-166.

Yarus, M. (1979) Prog. Nucl. Acid Res. Mol. Biol. 23, 195-225.

Youvan, D. C. & Hearst, J. E. (1982) Anal. Biochem. 119, 86-89.

Zwieb, C. & Brimacombe, R. (1980) Nucl. Acids Res. 8, 2397-2411.

Zwieb, C., Glotz, C. & Brimacombe, R. (1981) Nucl. Acids Res 9, 3621-3640.

Chapter Seven:

Role of am$^{\psi}$ in Eukaryotic mRNA Recognition

INTRODUCTION

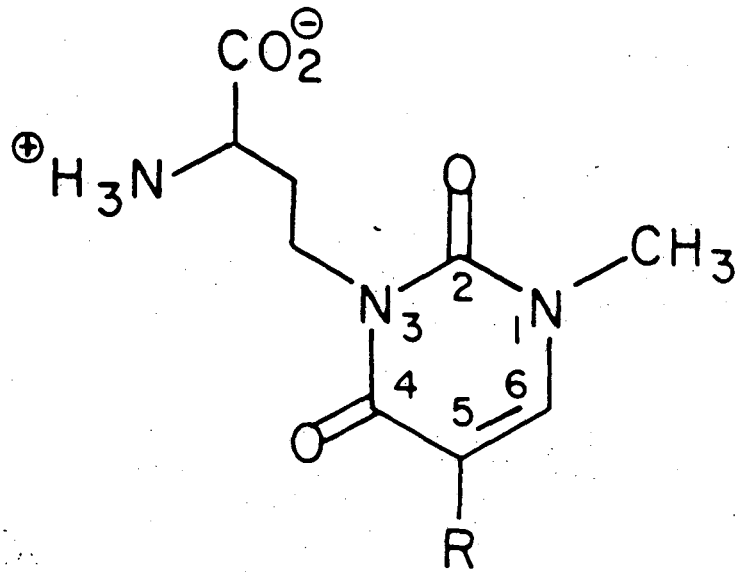
In Chapter 6, the structure and function of prokaryotic ribosomes were considered. While there are many similarities between prokaryotic and eukaryotic translation, striking differences are also present. Perhaps the most obvious difference is the manner of mRNA recognition. In prokaryotes, binding is regulated, at least to some extent, by base pairing of the 5' end of the mRNA to the 3' end of the 16S RNA (Shine & Dalgarno, 1975). This interaction does not exist in eukaryotes. Instead, there is a m^7G cap structure at the 5' end of the mRNA. This serves the double function of protecting the mRNA from exonuclease digestion and increasing levels of translation (reviewed by Filipowicz, 1978; Bannerjee, 1980). Speculation thus far on how the cap accomplishes increased translational efficiency has centered on the role of cap binding proteins. These proteins bind specifically to the cap structure and, with Mg^{2+} and ATP as cofactors, unwind the secondary structure of the mRNA (Sonenberg, 1981; Sonenberg et al., 1981). mRNAs with little or no secondary structure are translated with higher efficiency (Kozak, 1980).

While the presence of cap binding proteins can explain a large part of the increase in translation associated with m^7G, it is also clear there must be something else involved

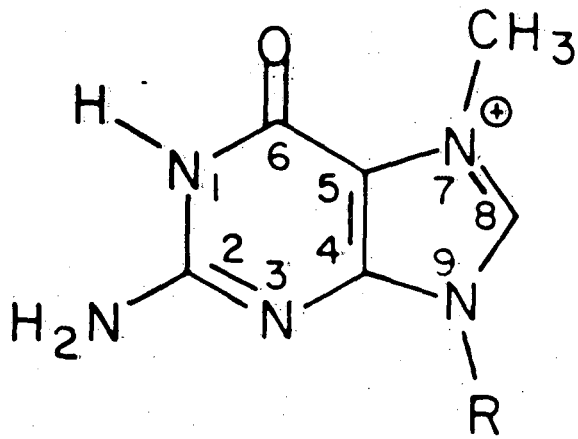
also. Kozak (1980) found that even mRNAs which had lost all secondary structure as a result of bisulfite reaction (causes complete conversion of cytidine to uridine) could have their translation inhibited by exogenously added m^7GDP. Therefore, the m^7G must have some role other than simply promoting the unwinding of secondary structure. Based on psoralen crosslinking results, we have proposed that the hypermodified base am (Figure 1) plays a major role in this recognition (Chapter 6). Inhibition of *in vitro* translation by am in a manner similar to that done with m^7G (Shafritz et al., 1976; Hickey et al., 1977; Adams et al., 1978; Hickey et al., 1976; Roman et al., 1976) was attempted to verify this proposal.

MATERIALS AND METHODS

The hypermodified nucleoside am was synthesized by Steve Isaacs of this laboratory in a procedure which will be fully described elsewhere with a complete structural characterization of the product. The synthesis started with protected pseudouridine (Sigma). After specific methylation at N1, the N3 position was alkylated with protected bromohomoserine followed by deprotection at all positions. The compound used in the translation studies still contained residual protecting groups. The partially deprotected products could be separated by reverse phase HPLC but were not characterized to determine which groups remained. *In vitro* translation was done with a rabbit reticulocyte lysate from



amψ



m^7G

Figure 1: Structure of m^7G and $am\psi$.

BRL and [^{3}H]leucine from Amersham. Each lysate contained the components provided by BRL as well as 0.15 μg globin mRNA and 5 μCi of [^{3}H]leucine and was incubated at 30° for 1 hr. Ten microliters of each sample was applied to Whatman GF/A glass filters, boiled in 10% trichloroacetic acid for 10 min, washed twice with 10% trichloroacetic acid, twice with 95% ethanol, and counted in a toluene scintillation fluid as described in Chapter 2.

RESULTS

The amount of translation observed using am$^\Psi$, pseudouridine, and uridine as inhibitors is shown in Figure 2. The am$^\Psi$ used has not been structurally characterized yet and is known to contain partially deprotected impurities so the results must be viewed as preliminary. The compounds which have been tested in Figure 2 are only the first of many that will be examined eventually. The concentration dependence of translation inhibition by am$^\Psi$ is similar to that observed for m^{7}GTP (Adams et al., 1978). Eventually, it will be possible to determine which parts of am are necessary for its activity. Clearly, it is not just simply the amino acid side chain because even $^\Psi$ gives some residual inhibition. Efforts are now underway to synthesize and test derivatives of am which have been only partially deblocked as well as analogs such as 1,3 dimethylpseudouridine. Presumably, the interaction is not simply between m^{7}G and am$^\Psi$ but also involves proteins which recognize specific aspects of both

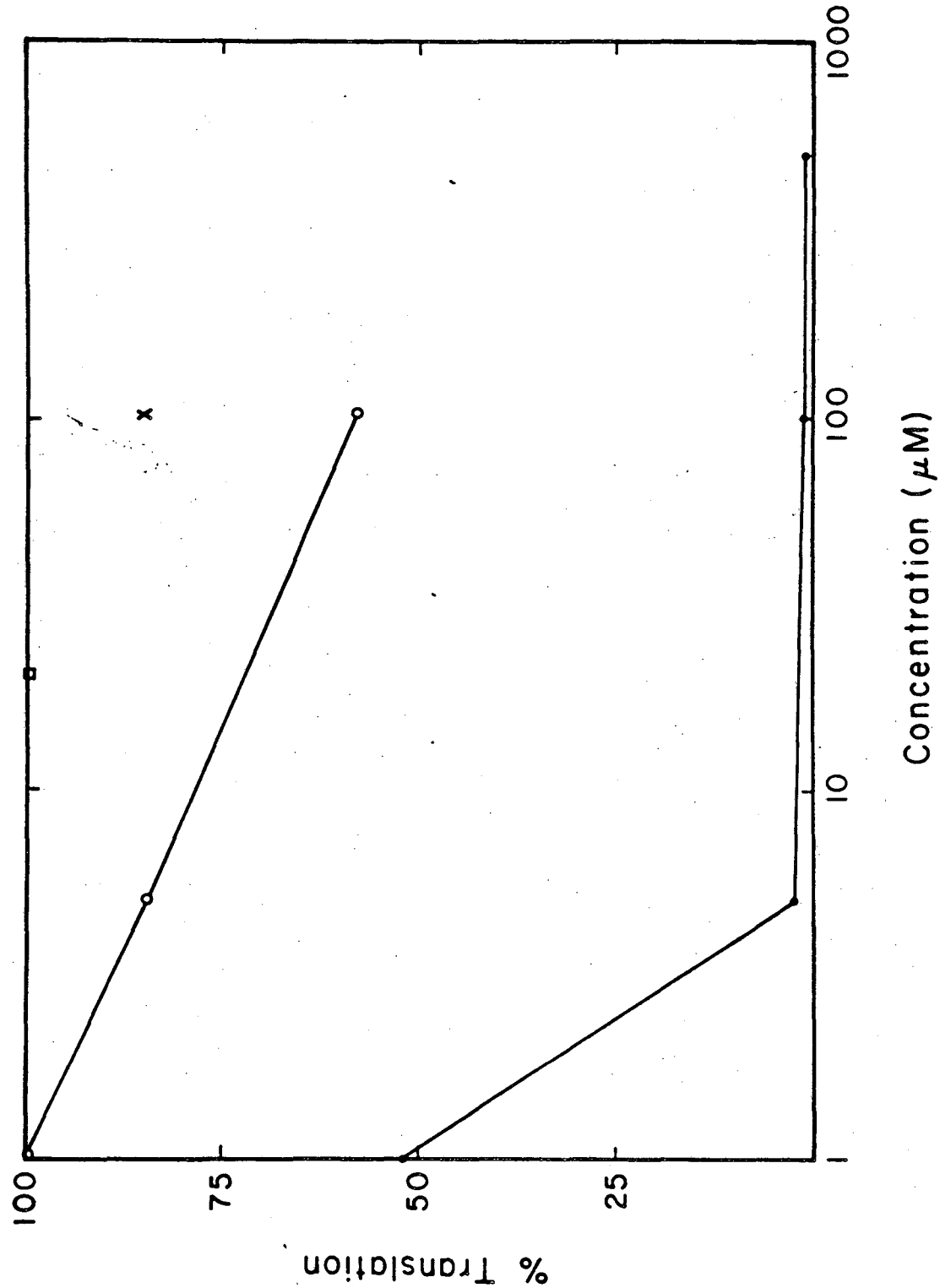

1000
100
10
1
Concentration (μM)
100
75
50
25
% Translation

Figure 2: Concentration dependence of translation inhibition. The amount of translation of globin mRNA is shown for samples to which various amounts of amΨ (•—•), Ψ (○—○), or uridine (□—□) has been added. These are relative to a control in which water was added. Samples were also run in the absence of added globin mRNA. This sample (x) was compared to a control to which water was added but no mRNA.

their structures.

The fact that am specifically interferes with initiation as opposed to elongation is shown by the sample in which addition of exogenous mRNA was omitted. The only translation which occurs is from polysomes which had already formed in the lysate before addition af am . In this system the concentration of endogenous mRNA is so low that reinitiation is unusual. Inhibition of translation by am is only 15% while in the normally initiating system, inhibition by am at the same concentration was over 95%.

DISCUSSION

While we have not yet been able to directly test which parts of am are important for interaction, some information can be gotten indirectly by examination of which parts of m^7G are important in translation inhibition (Roman et al., 1976; Hickey et al., 1977; Adams et al., 1978). These studies have used nucleotide analogs of m^7G to inhibit protein synthesis or mRNA binding. It was found that 2' or 3' phosphates on m^7G was of no importance while a 5' phosphate increased inhibition. A 5' di- or triphosphate was even better. The nature of the alkylating group at the 7 position was unimportant; only the presence of the positive charge mattered. The dependence of the inhibitory ability on the 5' phosphate has been attributed to the conformational rigidity imparted by the negative charge (Hickey et al., 1977). While this could play a role, it is also likely that the

amino group of am^7 could interact with the phosphate charge. When space filling models of the 5' end of the mRNA and the region around am^7 are made, it is clear that the least strained interactions are made with the β and γ phosphate oxygens. At the same time, the negative charge delocalized on the carboxylic acid group of am^7 interacts with the positive charge delocalized on the imidizole nitrogens.

Other analogs indicated that the 2' OH group and at least one of the N2 hydrogens were unimportant. Adams et al. (1978) also investigated other substituents for their effects but these results must be viewed with caution. Removal of the 2 amino group or replacement with a keto group has a large effect on the pK of the N1 hydrogen. The pK's of 7-methylguanine, 7-methylxanthine, and 7-methylhypoxanthine are 9.9, 8.9, and 7.5, respectively. While the pKs are not available for all the triphosphate compounds, they probably follow the same trend. The pK of m^7GTP is 7.1 so it is clearly sensitive to the pH in the range being studied. While m^7GDP will have significant amounts of both charged and uncharged N1 atoms, m^7XDP will have only small amounts of uncharged and m^7IDP will be virtually all charged. This is also the trend in inhibitory effects. Thus, this data cannot conclusively say whether it is the 2 amino group or the uncharged N1 position which is important. It is difficult to interpret the pH dependence of translation to determine which feature is important because both the m^7GTP and the m^7G cap will be similarly affected.

Space filling models of the two regions have been constructed to examine what other types of interactions might be present. In making this model, only the interactions contributed by the 5' end of the mRNA and the region around am$^{\Psi}$ were considered. In the ribosome, proteins will also certainly be involved and may displace some of the postulated interactions. In addition to the ionic interactions described above, the cytidine 5' to am$^{\Psi}$ may engage in normal Watson-Crick base pairing to m^{7}G. This would be allowed only if the N1 position is uncharged as proposed earlier. The presence of an N1 methyl group, which abolishes the inhibitory effect, would also prevent base pairing. An amino hydrogen from am$^{\Psi}$ can hydrogen bond to the imino nitrogen of the 3' cytidine while one of the cytidine's amino hydrogens can interact with a phosphate oxygen. At the same time, this cytidine can effectively stack over the nucleotide adjacent to m^{7}G. This base is generally a methylated purine which has a strong tendency to stack (Tazawa et al., 1980). The adenine two bases to the 3' side can stack on the opposite side while the adenine two bases to the 5' side of am$^{\Psi}$ can stack with m^{7}G. The region including the N1 methyl, C5, C6 and the ribose of am$^{\Psi}$ forms a relatively hydrophobic surface which would be ideally suited for contact with a protein. Proof of these, or other, interactions will have to await further experimentation.

The fact that the sequence surrounding am$^{\Psi}$ in eukaryotes is so highly conserved suggests a functional importance.

am$^{\Psi}$ has been definitively mapped in D. melanogaster, (Youvan & Hearst, 1981) and X. laevis (Salim & Maden, 1981) but the T_1 fragment in which it is located is highly conserved in all other eukaryotic species examined (Fuke and Busch, 1979). The secondary structure which brings it in contact with the 3' end of the 18S RNA is also highly conserved (Chapter 6). In prokaryotes, the 3' end of 16S is involved in mRNA recognition so it should be expected that eukaryotes would use a similar mechanism.

There are other, indirect data which support the m^7G-am$^{\Psi}$ interaction. There are three post-transcriptional modifications necessary to make am$^{\Psi}$. Pseudouridylation is an early, nuclear step followed by methylation at N1. Only after transport to the cytoplasm does attachment of the amino acid derivative occur (Brand et al., 1978). This means that the modification is not necessary for processing or transport and thus is probably functionally important. There is also a very similar hypermodified base present relatively frequently in prokaryotic tRNAs, acp^{3}U or X. This nucleoside differs only in that it is a uridine derivative rather than pseudouridine. In all prokaryotic tRNAs which have been found to contain X, the base immediately adjacent to it on the 5' side has always been m^7G (Sprinzl & Gauss, 1982). In the one case in which eukaryotic tRNA has been found to contain X, it was in a location where it could easily form a tertiary interaction with a m^7G (Chen & Roe, 1978).

While the presence of the am$^{\Psi}$-m^7G interaction does not

rule out the possibility of some base pairing recognition of mRNA to 18S RNA in eukaryotes, it certainly appears to obviate the need for it. Such base pairing schemes have fluorished so much principally because there is no good, uniform pairing present. Eukaryotes have simply evolved a different mechanism for mRNA recognition. There is apparently no way this interaction could be involved in complex regulation of translation but the strength of the Shine-Dalgarno pairing is not related in an obvious way to translational regulation either.

REFERENCES

Adams, B. L., Morgan, M., Muthukrishnan, S., Hecht, S. M., & Shatkin, A. J. (1978) J. Biol. Chem. 253, 2589-2595.

Bannerjee, A. K. (1980) Microbiol. Rev. 44, 175-205.

Brand, R. C., Klootwijk, J., Planta, R. J., & Maden, B. E. H. (1978) Biochem. J. 169, 71-77.

Chen, E. Y. & Roe, B. A. (1978) Biochem. Biophys. Res. Comm. 82, 235-246.

Filipowicz, W. (1978) FEBS Let. 96, 1-11.

Hickey, E. D., Weber, L. A., & Baglioni, C. (1976) Proc. Nat Acad. Sci. USA 73, 19-23.

Hickey, E. D., Weber, L. A., Baglioni, C., Kim, C. H., & Sarma, R. H. (1977) J. Mol. Biol. 109, 173-183.

Kozak, M. (1980) Cell 19, 79-90.

Roman, R., Brooker, J. D., Seal, S. N., & Marcus, A. (1976) Nature 260, 359-360.

Salim, M. & Maden, B. E. H. (1981) Nature 291, 205-208.

Shafritz, D. A., Weinstein, J. A., Safer, B., Merrick, W. C. Weber, L. A., Hickey, E. D., & Baglioni, C. (1976) Nature 261, 291-294.

Shine, J. & Dalgarno, L. (1975) Nature 254, 34-38.

Sonenberg, N. (1981) Nucl. Acids Res. 9, 1643-1656.

Sonenberg, N., Guertin, D., Cleveland, D., Trachsel, H. (1981) Cell 27, 563-572.

Sprinzl, M. & Gauss, D. H. (1982) Nucl. Acids Res. 10, r1-r55.

Youvan, D. C. & Hearst, J. E. (1981) Nucl. Acids Res. 9,

APPENDIX

While the techniques described in Chapter 6 were very successful in analyzing the structure of *E. coli* 16S RNA, meaningful results could not be obtained in two very specific problems in the structure of *D. melanogaster* 18S RNA. The methods used in these experiments are described below because both problems are potentially solvable with further work.

Crosslinking of *D. melanogaster* 5S and 18S RNA in solution can be done in high yield. If 18S RNA is in excess, greater than 50% of added 5S can be crosslinked. The two RNAs can be purified as described in Chapters 5 and 6. Both hybridization and crosslinking can be done in TMA I buffer, but higher yields are obtained in 0.15 M NaCl, 10 mM EDTA, 10 mM Tris (pH 7.5). Annealing is carried out at 60° for 3 hr. at RNA concentrations greater than 10 OD/ml. For crosslinking, the RNA is diluted to 5 OD/ml and 20 µg/ml HMT added. Lower yields are found with AMT, even at 1 mg/ml. Crosslinking is done at 5-10° for 15 min. The level of intermolecular crosslinking can be monitored by using ^{32}P labelled 5S RNA and running on a 4% polyacrylamide, 7 M urea gel.

Unfortunately, crosslinking of total 18S and 5S RNA yields too complex a pattern using the standard 2D gel system. Most of the off diagonal spots arise from intramolecular 18S crosslinks. The ideal way to avoid this problem is to place adducts on the 5S RNA which can be driven to

crosslink in the presence of 18S RNA. This was attempted with IMT and AZT, both of which can be reacted with RNA independent of 360 nm light (Chapter 4). Levels of incorporation were simply too low to get results. Use of 390 nm light to produce exclusively monoadducts would seem to be a good solution to this. A slightly different approach, using ^{32}P labelled 5S, was also unsuccessful because of the low specific activity attainable for general labelled RNA.

Site specific crosslinking of D. melanogaster 18S RNA was also attempted by affinity labelling of am with a psoralen derivative. AMT was coupled to an N-hydroxysuccinimide ester (S. Isaacs, unpublished results). This was stored at -20° in dioxane until needed. A 100-fold molar excess of the ester was added to 18S RNA in 50 mM triethanolamine (pH 8.1) in H_2O. Addition was done at room temperature over 6 hr with constant stirring. Concentrations were such that the final solution was 50% H_2O, 50% dioxane. After reaction an equal volume of .4 M NaCl was added followed by a 2x volume of ethanol. Precipitation was done at -20° overnight. After precipitation, the RNA was redissolved in .2 M sodium phosphate (pH 3.5), phenol extracted twice, and ethanol precipitated twice. The RNA was redissolved in TMA I, incubated at 37° for 30 min, and crosslinked at 5-10° for 20 min. The RNA was run hot on a 4% polyacrylamide, 7 M urea gel. The gel was stained with ethidium bromide and the RNA visualized with 375 nm light. Most of the RNA migrated like normal 18S RNA but 3 additional bands near the top of the gel

could also be seen. These were cut out and eluted electrophoretically. Each band was analyzed as in Chapter 6. Several off diagonal spots were observed. After reversal, 2-4 bands per spot were present. The psoralen derivative used was trifunctional so up to 3 bands should be expected. Unfortunately, there were not enough counts in the bands to analyze.

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720