THE 3'-END OF TURNIP YELLOW MOSAIC VIRUS RNA; APPLICATION OF NOVEL SEQUENCING TECHNIQUES

by

MELVIN SILBERKLANG

B.A., Columbia University

1971

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1976

(i.e. February, 1977)

Signature of Author Signature redacted

Department of Biology
September 24, 1976

Certified by Signature redacted

Thesis Supervisor

Accepted by Signature redacted

Chairman, Departmental Committee

ARCHIVES

MASS. INST. TECH.
FEB 1 1977
LIBRARIES

ABSTRACT

THE 3'-END OF TURNIP YELLOW MOSAIC VIRUS RNA; APPLICATION OF NOVEL SEQUENCING TECHNIQUES

by

Melvin Silberklang

Submitted to the Department of Biology on September 24, 1976 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

The 2×10^6 dalton MW RNA of Turnip Yellow Mosaic Virus (TYMV) can be aminoacylated *in vitro* with valine and is recognized by many tRNA-specific proteins; the aminoacylated RNA can also donate valine to polypeptides in a cell-free protein synthesis system. This evidence suggests the existence of a tRNA-like structure at the 3'-end of the viral RNA. In order to examine this possibility directly, the susceptibility of TYMV RNA to cleavage by partially purified preparations of the *E. coli* tRNA maturation nuclease, "RNase P", was exploited to allow preparation of a specific 4.5 S 3'-terminal RNA fragment. This fragment was purified from non-radioactive viral RNA, and its primary nucleotide sequence was investigated. Sequence information was derived entirely by *in vitro* [^{32}P] end group labeling techniques.

To facilitate primary sequence analysis of the TYMV RNA fragment, modifications were introduced into end group labeling and sequencing procedures previously developed in our laboratory. These included the adaptation of the two-dimensional homochromatography system to the analysis of partial nuclease digests of end group labeled RNA. In addition, new methods were developed for the use of polynucleotide kinase in nucleotide composition microanalysis and for the use of endonuclease P1 (from *Penicillium citrinum*) in the generation of partial digests of [^{32}P] end group labeled RNA for direct sequence analysis by two-dimensional homochromatography. The latter two procedures have also been shown to be applicable to DNA sequence analysis.

The 4.5 S TYMV RNA 3'-terminal fragment has been found to be 112 ± 3 nucleotides in length and to contain no modified bases. All the oligonucleotides present in T_1 and pancreatic RNase "fingerprints" of the fragment have been sequenced; the fragment does not contain the G-U-U-C-purine sequence typical

of loop IV of all tRNAs active in polypeptide chain elongation. Through the use of nuclease P_1, the sequence of 26 nucleotides from the 5'-end and 16 nucleotides from the 3'-end of the fragment has been deduced. An additional four nucleotides at the 5'-end were sequenced from a slightly longer 3'-terminal fragment contaminating one of our "RNase P" digests of TYMV RNA. The sequence at the 5'-end can be read, in one phase, as representing codons for serine and threonine followed by the termination codon U-A-A. As the published carboxy terminus of the TYMV coat protein is, in fact, Ser-Thr, this suggests that the 5'-end of the fragment includes the end of the TYMV coat protein gene, and the tRNA-like structure follows immediately after the termination codon. This has been corroborated by independent sequencing work from the laboratory of L. Hirth and collaborators (personal communication; Eur. J. Biochem. (1976), submitted). Furthermore, by comparison of data from our two laboratories sufficient oligonucleotide sequence overlaps were found to allow determination of a complete tentative 3'-terminal primary sequence of 115 nucleotides. This sequence can be loosely fitted into a tRNA-like coloverleaf secondary structure. However, further analysis of this structure awaits determination of a totally unambiguous primary sequence.

The sequencing methods developed during the course of this work have also been used to determine the first sequence of a eukaryotic organellar tRNA, that of Euglena gracilis $tRNA^{Phe}$, and have been used to analyze the sites of modification of 5S RNA treated in situ, in bacterial ribosomes, with monoperphthalic acid.

Thesis Supervisor: Dr. U.L. RajBhandary, Professor of Biochemistry

AUTOBIOGRAPHICAL NOTE

Date of birth: January 7, 1950

Place of birth: Augsburg, W. Germany

Education:

1967-1971 Columbia University
New York, New York

1971-1976 Massachusetts Institute of Technology
Cambridge, Massachusetts

Publications:

Chang, S.H., Hecker, L.I., Silberklang, M., Brum, C.K., Barnett, W.E. and RajBhandary, U.L. (1976)
Nucleotide Sequence of Phenylalanine tRNA from Chloroplasts of Euglena gracilis. Fed. Proc. 35, 150.

Silberklang, M., Prochiantz, A., Haenni, A.L. and RajBhandary, U.L. (1976)
Novel Sequencing Techniques Applied to the 'tRNA-like' 3'-Terminus of Turnip Yellow Mosaic Virus RNA. in Animal Virology, vol. IV, ICN/UCLA Symposia on Molecular and Cellular Biology, ed. Baltimore, D. and Huang, A.S. (Academic Press, N.Y.), in press.

Chang, S.H., Brum, C.K., Silberklang, M., RajBhandary, U.L., Hecker, L.I. and Barnett, W.E. (1976)
The First Nucleotide Sequence of an Organelle Transfer RNA: Chloroplastic $tRNA^{Phe}$. Cell, in press.

Silberklang, M., Prochiantz, A., Haenni, A.L. and RajBhandary, U.L. (1976)
Studies on the Sequence of the 3'-Terminal Region of Turnip Yellow Mosaic Virus RNA. Eur. J. Biochem., submitted.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor U.L. RajBhandary for his instruction, guidance and stimulating intellectual influence. Many elegant experiments were inspired during our candid dialogues.

I am also deeply indebted to my colleagues, Amanda Gillum, Alain Prochiantz and Simon Chang, with whose collaboration this work was completed. The encouragement and collaboration of Anne-Lise Haenni is also acknowledged.

I am very grateful to Professor H.G. Khorana and many members of his research group for very valuable discussions and for the use of materials and facilities which were vital for the completion of this work.

I would also like to thank the members of my research group, past and present, for their help, suggestions and valued friendships.

Finally, my sincerest thanks to June Stillman and Liza Bingham for excellent typing and expert editorial assistance, and to Joyce Heckman for help with the figures; and thanks to Roselyn for waiting.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A. Proteins

Pan RNase	bovine pancreatic RNase A
SVP	snake venom phosphodiesterase
SP	bovine spleen phosphodiesterase
BAP	bacterial (*E. coli*) alklaine phosphatase
CAP	calf intestinal alkaline phosphatase
ValRS	valyl-tRNA synthetase
HisRS	histidyl-tRNA synthetase
BSA	bovine serum albumin

B. Nucleic Acids and Components

$tRNA_f^{Met}$	formylatable initiator methionyl tRNA, prokaryotic
$tRNA_i^{Met}$	initiator methionyl tRNA, eukaryotic
D	dihydrouridine
T	ribothymidine
Ψ	pseudouridine
m^1A	1-methyl adenosine
m^1G	1-methyl guanosine
m^7G	N^7-methyl guanosine
Nm	2'-O-methyl ribonucleoside
s^4U	4-thiouridine
$4(abu)^3U$	N^3-(4(2-amino)butyryl)uridine
ms^2i^6A	2-methylthio-N^6-isopentenyl adenosine
dA^{Bz}	N^6-benzoyl deoxyadenosine
dC^{An}	N^4-anisoyl deoxycytidine
dG^{iB}	N^2-isobutyril deoxyguanosine

C. Chemicals and Chromatographic Supplies

DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
NTA	nitrilotriacetic acid
PGA	phosphoglyceric acid
TEA	triethylamine
TEAB	triethylammonium bicarbonate
Tris	tris(hydroxymethyl) aminomethane
BD-cellulose	benzoylated-DEAE-cellulose
DEAE-cellulose	diethylaminoethyl cellulose
DEAE-paper	DEAE-cellulose paper
DEAE-Sephadex	diethylaminoethyl sephadex
PEI-cellulose	polyethyleneimine-impregnated cellulose
RPC-5	reversed phase chromatography resin-5

D. Viruses

AMV	Avian Myeloblastosis Virus or Alfalfa Mosaic Virus
BMV	Brome Mosaic Virus
CMB	Cucumber Mosaic Virus
EMC	Encephalomyocarditis Virus
EMV	Eggplant Mosaic Virus
MLV	Murine Leukemia Virus
OMV	Okra Mosaic Virus
TMV	Tobacco Mosaic Virus
TRV	Tobacco Rattle Virus
TYMV	Turnip Yellow Mosaic Virus
WCMV	Wild Cucumber Mosaic Virus

E. General

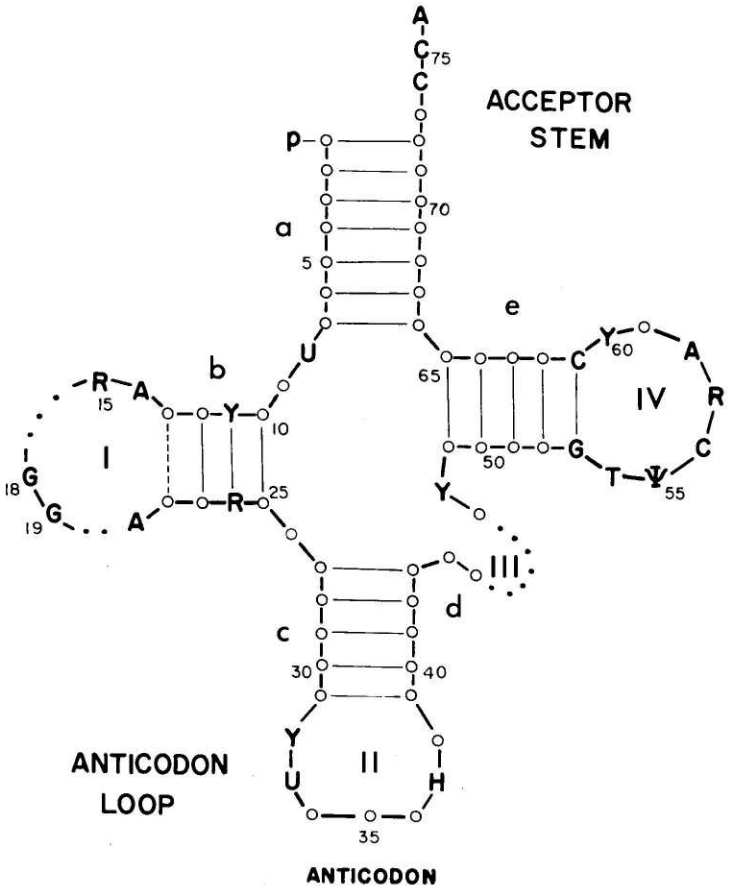
TLC	thin layer chromatography

INTRODUCTION

The major biological function of transfer RNA is in the translation of the sequence of messenger RNA trinucleotide codons into an amino acid sequence. tRNA plays the role of "adapter" (1) on the ribosome, base-pairing its anticodon end directly with messenger RNA while carrying, at its other end, the growing polypeptide chain. For this central role in the processing of genetic information, tRNA has evolved as a class of molecules of similar size with similar highly ordered secondary and tertiary conformations. The general nature of the secondary structure of tRNA was first deduced by Holley (2), who proposed the now familiar "cloverleaf" model (Fig. 1); all tRNAs sequenced to date can be fitted into such a structure (3). More recently, the precise tertiary conformation of yeast $tRNA^{Phe}$ has been deduced from 2.5 Å resolution X-ray diffraction studies of both the orthorhombic (4,5,6) and monoclinic (7,8) crystal forms of the molecule. This tertiary structure model now serves, like the cloverleaf pattern before it, as a point of reference in interpreting the properties of all tRNAs, either in crystalline arrays or in solution.

All tRNAs have the common 3'-terminal nucleotide sequence C-C-A, the 2' or 3' hydroxyl (9,10) of the terminal adenosine being the site of aminoacylation. Additional invariant and "semi-invariant" features of known tRNA sequences are summarized in the generalized cloverleaf model of Fig. 1,

Figure 1. The cloverleaf pattern: generalized secondary structure for tRNA. The position of "invariant" and "semi-invariant" nucleotides is shown. The numbering system is that of yeast $tRNA^{Phe}$. Y stands for pyrimidine, R for purine, H for a hypermodified purine. R_{15} and Y_{48} are usually complementary. Other constant features are noted in the text.


A
C
C
75
ACCEPTOR
STEM
p
a
5
70
e
b
U
R
A
15
Y
10
I
G
18
G
19
A
R
25
65
C
Y
60
A
IV
R
C
G
T
Ψ
55
50
Y
III
d
c
30
40
Y
U
II
H
35
ANTICODON
LOOP
ANTICODON

which also indicates the standard nomenclature(s) for the cloverleaf helical, base-paired "stems" and single-stranded "loops". tRNAs range in size from about 75 to 93 nucleotides, the extra nucleotides of the longer molecules being accommodated in stem d and loop III, the so-called "variable loop". In general, structural variability in tRNA is restricted to stems b and d and loops I and III.

Stems a and e usually contain seven and five base-paired nucleotides respectively; among the few exceptions to these rules are the histidine tRNAs of S. typhimurium and E. coli (11), which have eight base-paired nucleotides in stem a, and most prokaryotic translation initiator methionine tRNAs, which have only six base-paired nucleotides (the seventh nucleotide is present, but is non-pairing) in this stem (3,11). Stem c and loop II make up the anticodon arm of tRNA (Fig. 1). The tRNA cloverleaf structure is generally written so as to accommodate only five nucleotides in stem c, although the primary nucleotide sequence of some tRNAs allows for a potential additional A·U or A·Ψ base pair between the first and last nucleoside of loop II (eg., ref. 11: pp. 18, 19, 36, 40, 45). Staphylococcal glycine tRNAs involved in cell wall biosynthesis (12) also violate this rule, having six base-paired nucleotides in stem c (11). An interesting example of the importance, in tRNA structure, of having seven nucleotides in loop II is provided by a mutant version of S. typhimurium $tRNA_1^{Gly}$, the Suf D mutation, which has an

extra C in, and hence eight nucleotides in loop II (13). This tRNA behaves as if it had a tetranucleotide anticodon, and can therefore act as a frameshift suppressor -- direct evidence that a seven membered loop II is essential to maintaining the proper codon reading frame during protein synthesis. Loop IV is always constant, having seven nucleotides.

In addition to invariant structural features, there are also various invariant and "semi-invariant" features of nucleotide sequence in tRNA (Fig. 1), most of which are involved in tertiary structural interactions (3). For example, there are two invariant G residues, G_{18} and G_{19}, in loop I, which are involved in hydrogen bonding interactions with loop IV. The invariant nucleotide U_8 is hydrogen bonded to A_{14}, and the invariant pyrimidine residue at position 48, which is usually modified, is complementary to $purine_{15}$, with which it interacts. Also worthy of note is the sequence G-T-Ψ-C-purine (or other derivatives of G-U-U-C-purine) of loop IV, which is present in all tRNAs, prokaryotic and eukaryotic, that are active in polypeptide chain elongation; it is missing only in eukaryotic cytoplasmic initiator methionine tRNAs (11, 14-20) and the Staphylococcal cell wall biosynthesis $tRNA^{Gly}$s. The T and Ψ in this sequence also exemplify nucleoside modifications that are "semi-invariant" in tRNA. The sequence G-T-Ψ-C has been implicated as being involved in a base-pairing interaction with 5S RNA in bacterial ribosomes (161; vide infra, Results IIIB).

Such extensive constant features, shared by all tRNAs, might be due both to evolution from a common ancestral proto-tRNA species (21), or group of such species (22), and to the common selective constraint of essential functional interaction with ribosomes and various proteins. The necessity of decoding information for 20 different amino acids dictates that all independent organisms must have at least 20 distinct species of tRNA. In addition, the degeneracy of the genetic code implies that most organisms will also have multiple iso-accepting tRNA species for certain amino acids to allow translation, by codon-anticodon base-pairing, of all possible codons for those amino acids. Thus, each organism has available, in its complement of tRNAs, a class of molecules possessing very similar, but clearly distinguishable structures, each of which can be acylated with one specific amino acid. A plausible argument could be made that use of these tRNA molecules for purposes other than protein synthesis reflects both the efficiency and opportunism of nature.

The range of known tRNA functions outside the domain of ribosomal protein synthesis is considerable. Uncharged tRNA in the ribosomal acceptor site, for example, promotes synthesis of ppGpp and pppGpp by "stringent factor" in rel^+ E. coli (23); tRNA thus mediates the "stringent" response of these strains to starvation for an essential amino acid, wherein G tetra- and penta-phosphate accumulate and inhibit synthesis of stable RNA species. Aminoacyl tRNA functions

as corepressor, in some cases, of cognate amino acid biosynthetic operons in bacteria (24). Aminoacyl tRNA also participates as amino acid donor in the reactions catalyzed by the various aminoacyl tRNA transferase enzymes (aminoacyl-tRNA protein transferases, aminoacyl-tRNA phosphatidyl glycerol transferases, and aminoacyl-tRNA N-acetylmuramyl-pentapeptide transferases) (25). tRNA may also function as an allosteric effector of enzyme activity. The work of Goebel and Helinski (26), for example, suggests that *in vitro* inhibition of the activity of the *E. coli* enzyme endonuclease I by tRNA (27) may involve modification of the activity of the enzyme from double strand scission of DNA to a low level of single strand nicking; the enzyme may also be associated with tRNA *in vivo* (M. Silberklang, unpublished observation). In another instance, it has been suggested that *Drosophila* $tRNA^{Tyr}$ may act as a specific inhibitor of the enzyme tryptophan pyrrolase in that organism (28); however, this model was recently challenged (29).

Participation of tRNA in viral development has attracted considerable interest. The host-virus relationship is such that the virus must rely on the host cell for a considerable proportion of the transcriptional and translational apparatus necessary for expression of the viral genome. This generally holds true for the complement of tRNA available for protein synthesis. T4 bacteriophage, however, is an exception; during infection of *E. coli*, this phage induces the synthesis

of at least eight new tRNAs (30). The same tRNA species can be detected in the products of *in vitro* transcription of T4 DNA (31). It has been reported (32) that the T4-coded tRNAs read certain codons for which host tRNAs are rare. It has also been found (33) that in phage deletion mutants lacking these tRNA genes, the *in vivo* rate of synthesis of tail fiber proteins P34 and P37 is reduced by 40-50 percent, and this has been correlated with a similar reduction in burst size. Moreover, in a particular *E. coli* strain isolated from a hospital patient, T4 tRNA gene deletion mutants would not grow at all (33). Taken together, these data indicate that a major function of the phage-coded tRNAs is to insure optimal rates of translation in the maximum mumber of hosts by supplementing the reading capacity for certain codons used more commonly in the viral than in the host genome. Possibly also related to this is the finding that T-even phages induce the specific nucleolytic cleavage of host $tRNA_1^{Leu}$ upon infection (34). The resulting reduction in the host tRNA reading capacity for the leucine codon CUG may contribute to the cessation of host protein synthesis (35); presumably, the phage genome uses this codon less frequently than that of the host (34,35). Phage T4 infection also induces a structural modification in host valyl-tRNA synthetase by addition of a phage-coded subunit (36), which leads to tighter binding of $tRNA^{Val}$ (37) and to the capacity to also aminoacylate yeast tRNA (36); the functional signifi-

cance of this alteration, however, is not clear (38).

E. coli phage T5, too, codes for at least 14 different tRNAs (39). The phage T5-coded tRNAs include two methionine tRNAs, one of which can be formylated in vitro (i.e., it behaves like a translational initiator tRNA) (40). As the single methionine codon, AUG, is recognized equally well by host tRNA, these T5 tRNAs, at least, may have functions other than increasing the codon reading capacity of the host tRNA pool. A similar conclusion about T5 tRNAs has been reached by Henckes et al. (41), who found that T5 also coded for two isoaccepting $tRNA^{Ser}$s that read the serine codons UCC and UCU, respectively. These codons were also efficiently read by the tRNAs of the host strain of E. coli. The possibility remains, however, that these $tRNA^{Ser}$s are useful in supplementing the reading capacity of the tRNA of other host strains found in nature.

Eukaryotic DNA viruses may also contain tRNA genes. Although early reports of Herpes virus coded arginine tRNA (42) were challenged (43) and later retracted (44), it has been reported, for example, that phenylalanine acceptor activity goes up considerably in tRNA from HSV-2 infected cells (45). This is a phenomenon that requires further investigation.

tRNA also plays a special role in the life cycle of RNA viruses. It was first observed that purified Avian Myeloblastosis Virus (AMV) virions yielded a fraction of RNA,

soluble in 2 M NaCl, that could be aminoacylated in vitro (46). This RNA is methylated to the same extent as host tRNA (47); it hybridizes equally to the DNA of normal and transformed cells, and can compete with host tRNA for hybridization sites (48). Although the tRNA is cellular in origin, the selection of tRNA species incorporated into the virion, totalling about 200 tRNA molecules per 70S viral RNA (49), seems to be specific; the ratios of tRNA species in the virions differs considerably from that in the cell (46,49), and is also different for different RNA tumor viruses (50).

Most of these tRNA species are easily separable from the 70S viral RNA by sucrose density gradient centrifugation (51). Their role in viral development is not known. Another fraction of the tRNAs is associated with the 70S RNA, and is separable only after heat denaturation. Two components of the associated tRNA fraction can be separated by sequential heating in EDTA at 60°C and 80°C (52). Heating at 60-70°C separates the 70S viral RNA into 35S subunits and releases about 4 tRNAs per 35S subunit (53). It has been speculated that these may act as linkers, or otherwise, to hold together the 35S RNA subunits in the virion (54), although more recent evidence indicates that this fraction of the associated tRNA is almost as complex as the non-associated tRNA fraction (53). Further heating at 80°C, however, releases a specific tRNA, which has been shown to be host

$tRNA^{Trp}$ for Rous Sarcoma Virus (RSV) (55) and AMV (56) and host $tRNA^{Pro}$ for Murine Leukemia Virus (MLV) (57); this tRNA has been shown to be the primer for viral reverse transcriptase (58,59,55). The primer tRNA can apparently both hybridize with the viral 35S RNA (56) and bind tightly to the free reverse transcriptase enzyme (60). An interesting and unusual common feature of these primer tRNAs is that they contain the sequence Ψ-Ψ-C-G in loop IV instead of T-Ψ-C-G (Fig. 1) (55,61).

Another interesting instance of the role of tRNA in host-viral relationships is the now well-established presence of tRNA-like conformations at the 3'-termini of many plant viral RNAs (62). This phenomenon was first studied in Turnip Yellow Mosaic Virus (TYMV), the RNA genome of which was shown to be enzymatically aminoacylatable with valine (63). It was demonstrated (64) that the reaction was specific for valine, both with E. coli and with yeast aminoacyl-tRNA synthetase extracts; the reaction was dependent on ATP and was kinetically similar to the enzymatic aminoacylation of $tRNA^{Val}$. The pH stability of TYMV valyl-RNA was found to be similar to that of valyl-$tRNA^{Val}$ at 37°C, pH 8.6; the amino acid was released with a half-life of 60 min. Attached [^{14}C]-valine co-sedimented with intact TYMV RNA upon sucrose density gradient centrifugation. In further studies (65), it was demonstrated that TYMV valyl-RNA could be chemically acetylated, using conditions that acetylate the α-amino group

of aminoacyl tRNA; furthermore, using this acetylvalyl-RNA, it was observed that the enzyme peptidyl tRNA hydrolase (66) releases acetylvaline, as it does with acetylvalyl-tRNA. Also, when TYMV valyl-RNA was subjected to digestion with pancreatic RNase or snake venom phosphodiesterase, the amino acid was released as valyl-adenosine and valyl-AMP, respectively. When the TYMV valyl-RNA was digested with T_1 RNase, the unique valyl-oligonucleotide released was different from that resulting from digestion of host (cabbage) valyl-$tRNA^{Val}$. It was thus clearly demonstrated that valine can be enzymatically attached to the 3'-end of the TYMV RNA genome by an ester bond. Subsequently, it was shown that the aminoacylation of TYMV RNA could be performed with highly purified E. coli valyl-tRNA synthetase, but only after prior treatment with tRNA nucleotidyl transferase and ATP (67). This implies that the 3'-terminus of the viral RNA is ...C-C_{OH}, which can be enzymatically converted to ...C-C-A_{OH}, as in tRNA. The 3'-terminal adenosine, then, is the site of aminoacylation.

In addition to the enzymes so far discussed, valyl-tRNA synthetase, tRNA nucleotidyl transferase and peptidyl-tRNA hydrolase, TYMV RNA is also recognized by the E. coli tRNA maturation endonuclease, "RNase P"[1], which releases a 4.5 S 3'-terminal RNA fragment (68). Furthermore, TYMV valyl-RNA

---

[1]The "RNase P" used in this work was a preparation purified up to the DEAE-Sephadex chromatography step of Robertson, Altman and Smith (79). The possibility that TYMV RNA is cleaved by enzyme(s) other than "RNase P" cannot, therefore, be ruled out (see Discussion ID).

can form a ternary complex with GTP and protein elongation factor EF-I (wheat germ) (69) or factor EF-T (E. coli) (70), and can donate valine to polypeptides in an E. coli cell-free protein synthesis system (71). Taken together, this evidence strongly implies the existence of a tRNA-like structure at the 3'-end of TYMV RNA.

There is also considerable evidence for a tRNA-like structure at the 3'-end of other plant viral RNAs (review: 62). In the TYMV group (Tymovirus), Okra Mosaic Virus (OMV) RNA, Egglant Mosaic Virus (EMV) RNA, and Wild Cucumber Mosaic Virus (WCMV) RNA have been reported to be enzymatically acylatable with valine[2] (72,73). The groups Bromovirus (74, 75) and Cucumovirus (75) can be aminoacylated with tyrosine and Tobacco Mosaic Virus (TMV) with histidine (69,76). TMV RNA is also recognized by tRNA nucleotidyl transferase (77), can be methylated by E. coli ribothymidine methylase (78), and, when aminoacylated with histidine, will form a ternary complex with GTP and wheat germ EF-I (69).

It has been shown that, unlike aminoacylated TYMV RNA, aminoacylated BMV RNA cannot donate tyrosine to polypeptides in a cell-free protein synthesis system (81). Also, when the 3'-terminal nucleoside of BMV RNA is oxidized with sodium periodate to generate the dialdehyde derivative of its ribose moiety (the ribose moiety of the 5'-"cap" m^7G is presumably also oxidized), the RNA continues to function as a messenger

---

[2]Interestingly, EMV virions may also carry an associated host $tRNA^{Lys}$ (72,80).

for *in vitro* protein synthesis (82); hence, aminoacylation must not be essential for messenger activity. Although there has been some speculation on the matter, the function of the 3'-terminal structures in plant viral RNAs remains unknown (see Discussion).

In experiments related to those described above, it has been found that other viral RNAs have internal tRNA-like conformations that may be revealed by partial nuclease digestion. The first such report claimed that partially enzyme-digested TMV RNA could be aminoacylated with serine and methionine (83); however, contamination of the crude synthetase enzyme preparations and/or of the TMV RNA preparations with host tRNA could not be rigorously excluded, and these results have not been confirmed (84). Recently, it has been shown that Mengo Virus RNA can be aminoacylated with histidine (85) and Encephalomyocarditis (EMC) Virus RNA with serine (86). Each of these viruses, however, is known to have a 3'-terminal poly(A) sequence (87,88); also, in the case of Mengo Virus, it has been shown that the aminoacylated RNA is not as large as the intact genome (85), though much larger than tRNA. It may therefore be assumed that the tRNA-like structure is internal in this RNA, and that the acceptor activity is exposed upon partial degradation by nucleases that contaminate the crude aminoacyl-tRNA synthetase preparation (89). Attempts to aminoacylate Polio Virus RNA, which also ends in a 3'-terminal poly(A) (90), have failed (76), as have attempts to

aminoacylate various phage RNAs (76,89). However, partially nuclease digested phage R17, MS2 and Qβ RNAs can act as substrates for the tRNA nucleotidyl transferase (89), and accept both C and A. These experiments (89) used commercially prepared snake venom phosphodiesterase to treat the RNA prior to reaction with tRNA nucleotidyl transferase; it was assumed that the fragmentation of the viral RNAs was caused by an endonuclease contaminating the commercial phosphodiesterase enzyme, while the phosphodiesterase partially removed the 3'-terminal C-C-A (11) of the viral RNA. As the endonuclease contaminating commercial snake venom phosphodiesterase does, in fact, have a pronounced preference for cleavage of the ...CpA... phosphodiester bond (J. Heckman, B. Baumstark and M. Silberklang, unpublished observations), it may have been a fortuitous choice of enzymes for the purpose of revealing internal C-C-A sequences. (We believe this nuclease to leave 3'-phosphomonoesters, so that the additional action of contaminating phosphatase must be invoked to explain nucleotidyl transferase susceptibility.) However, since the exact RNA features required for recognition by tRNA nucleotidyl transferase are not known, it may be an overstatement to call these bacteriophage RNA internal C and A accepting sequences "tRNA-like".

A further example of an internal tRNA-like structure in a viral genome is found in the "species I" RNA transcribed in vivo from T-even phage DNA (91). The primary nucleotide

sequence of this RNA shows it to contain a region that closely resembles a tRNA, but with a 19 nucleotide insertion into loop I of the cloverleaf structure (92). No function for this transcription product is known. tRNA-like primary nucleotide sequence features were also found in the 156-nucleotide, stable, VA RNA species (93) induced by Adenovirus-2 infection of KB cells (94).

The fact that the TYMV RNA 3'-end, in particular, had been shown to be recognized even by the protein synthetic apparatus, led us to choose this RNA for the studies described herein. In collaboration with A. Prochiantz and A.L. Haenni of the University of Paris, we decided to examine directly the tRNA-like characteristics of this region of the viral RNA by determining its primary nucleotide sequence. We set out to determine how closely the viral RNA mimicked such constant tRNA features as the presence of the sequence G-U-U-C-purine, or of modified bases, and whether we could identify specific features that allow recognition by valyl-tRNA synthetase. The susceptibility of TYMV RNA to specific cleavage by E. coli "RNase P" (68) made possible the preparation of a near-homogeneous 3'-terminal RNA fragment for sequence analysis (work of our collaborators in Paris).

In the course of this work, reports have appeared on partial 3'-terminal nucleotide sequences of TMV RNA (95,96) and of EMV RNA (97); no tRNA-like cloverleaf conformations were suggested from the data obtained (vide infra, Discussion). Recently, a preliminary report has also appeared on the nearly

identical 3'-terminal nucleotide sequences of the four RNA components of BMV (98,233). These studies were all done with in vivo [^{32}P]-labeled RNA. However, with RNA species that must be recovered from plant or animal tissues, it is often difficult to label in vivo with [^{32}P] to sufficiently high specific activity for the application of classical radiochemical sequencing techniques (99). We chose to work with non-radioactive viral RNA, and apply the in vitro [^{32}P] end group labeling techniques first described by Szekely and Sanger (100), and further developed in our laboratory (18-20, 101,102).

SUMMARY OF THE CONTENTS OF THIS THESIS

In the course of our current work, several modifications have been introduced into our established methods for the use of [^{32}P] end group labeling in RNA sequence analysis; in addition, several new sequencing methods have been developed:

1. A sensitive method has been developed for modified nucleoside composition microanalysis; this involves the use of polynucleotide kinase to phosphorylate the 5'-hydroxyl groups of 3'-mononucleotides in tRNA hydrolysates, and subsequent fractionation of the [^{32}P]-labeled nucleoside 3',5'-diphosphates by two-dimensional thin layer chromatography. An additional treatment with nuclease P_1 (from Penicillium citrinum) can be used to convert the nucleoside diphosphates to nucleoside 5'-monophosphates; these fractionate with superior resolution on thin layer chromatography. Analysis of

nucleoside 5'-monophosphates also allows the general use of UV marker compounds for nucleotide identifications, as such compounds are more readily available in the monophosphate form.

2. It has proved possible to scale down our previously published procedures for tRNA sequence analysis (18,101) so that T_1 or pancreatic RNase digestion of the substrate, followed by [^{32}P] end group labeling of the resulting oligonucleotides, can be performed on as little as 0.5 μg of RNA.

3. By adaptation of the two-dimensional homochromatogrpahy system of analysis (103), it has proved possible to sequence 5'-[^{32}P]-labeled oligonucleotides accurately and efficiently. Longer sequences, which are beyond the resolution of the DEAE-paper ionophoretic system of analysis used in previous work from this laboratory (102), can also be analyzed routinely by two-dimensional homochromatography.

4. A method has been developed which allows the 5'-terminal or 3'-terminal regions of RNA molecules as large as viral genomes to be sequenced directly. This involves the use of nuclease P_1 (from Penicillium citrinum) to generate partial digests of 5'- or 3'-[^{32}P] end group labeled oligonucleotides or high molecular weight RNA for analysis by two-dimensional homochromatography. Preliminary results indicate that this procedure may also be used for DNA sequence analysis.

Application of these methods to the 3'-terminal RNA fragment prepared by "RNase P" cleavage of TYMV RNA (68) has

led to the following results. The 3'-terminal RNA fragment, which is 112±3 nucleotides long, contains no modified nucleosides. All the oligonucleotides resulting from complete T_1 or pancreatic RNase digestion of the fragment have been sequenced; no U-U-C-purine sequence was found. Through the use of nuclease P_1 on 5'-[^{32}P]-labeled TYMV RNA 3'-terminal fragment and on 3'-[^{32}P]-labeled intact TYMV RNA, the sequence of 26 nucleotides from the 5'-end and 16 nucleotides from the 3'-end of the fragment has been deduced. An additional four nucleotides at the 5'-end were sequenced from a slightly longer 3'-terminal fragment contaminating one of our "RNase P" digests of TYMV RNA. Based on this work, it has been possible to compare our results with those of L. Hirth and collaborators (104); from the combined data it is possible to order the rest of the oligonucleotides we have sequenced, and hence to propose a tentative primary sequence for the 3'-115 nucleotides of TYMV RNA. A possible tRNA-like cloverleaf secondary structure for this sequence and the possible overlap of this sequence with the terminal region of the TYMV coat protein cistron are discussed.

Further application of the sequencing methodologies we have developed has also allowed the determination of the first nucleotide sequence of a eukaryotic organellar tRNA, Euglena gracilis chloroplast $tRNA^{Phe}$, and preliminary determination of sites of monoperphthalic acid in situ modifications of 5S RNA in bacterial ribosomes.

MATERIALS AND METHODS

I. Materials

A. Enzymes

Ribonucleases T_1 and T_2 were from Sankyo Co., Ltd. (Tokyo, Japan). RNase T_1 was further purified by U. L. RajBhandary through the chromatographic steps of Rushizky and Sober (166); the peak column fraction, at 2.4 U/μl (ca. 1 mg/ml), was stored at -20°C. RNase T_2 was dissolved in water at 1 U/μl and stored at -20°C.

Pancreatic ribonuclease A (Pan RNase), snake venom phosphodiesterase (SVP), and spleen phosphodiesterase (SP) were obtained from Worthington Biochemicals Corp.. Pan RNase was dissolved in water at 0.125-10 mg/ml and stored at -20°C. SVP was dissolved in 50 mM tris-HCl, 5 mM potassium phosphate, pH 8.9, at 1 mg/ml and stored at 4°C; when stored at this concentration, there was no appreciable loss of activity for up to 2 months. SP was dissolved in 10 mM tris-HCl, pH 7.6, 10 mM $MgCl_2$, at 5 U/ml and stored at -20°C.

Bacterial alkaline phosphatase (BAP), calf intestinal alkaline phosphatase (CAP), and yeast hexokinase (HK) were supplied by Boehringer Mannheim Corp.. HK was also purchased from P-L Biochemicals, Inc.. Ammonium sulfate suspensions of BAP, CAP and HK were generally stored at 4°C. BAP was occasionally diluted with water to a concentration of 2.5 U/ml and stored at -20°C. HK was sometimes diluted to 2.1 U/ml in 50% glycerol, 5 mM sodium

acetate, pH 5.0, and stored at -20°C; under these conditions, no loss of activity was observed for up to 1 month.

Nuclease P_1 from Penicillium citrinum (105) was purchased from Yamasa Shoyu Co., Ltd. (Tokyo, Japan). The lyophilized powder was dissolved in 50 mM tris-maleate, pH 6.0, at 1 mg/ml and stored at -20°C. It was diluted, as necessary, in 50 mM ammonium acetate, pH 5.3, the buffer used for RNA and DNA digestion; at a concentration of 0.002 mg/ml in this buffer, it could be stored at -20°C for 2 weeks without loss of activity.

T4 polynucleotide kinase, prepared according to Panet et al. (106), was generously supplied by Dr. A. Gillum. Highly purified tRNA nucleotidyl transferase from E. coli was a kind gift of Dr. D. Carre to A. Prochiantz.

Yeast phosphoglycerate kinase (PGK) and rabbit muscle glyceraldehyde-3-phosphate dehydrogenase (GAP-DH), used in the preparation of γ [^{32}P] ATP, were purchased from Boehringer Mannheim Corp.. The ammonium sulphate suspensions were stored at 4°C. A crude mixture of E. coli nucleoside mono- and diphosphate kinases, prepared according to Bishop (107) and dialyzed free of ammonium sulphate, was a kind gift of Dr. T. Sekiya; it was stored at -20°C and was used in the preparation of α [^{32}P] ATP.

B. Chemicals

The disodium salt of ATP was purchased from Sigma Chemical Co.; it was dissolved in water, titrated to pH 7 with NaOH, and

adjusted to 50 mM spectrophotometrically, assuming a molar extinction coefficient of $\mathcal{E}^{260} = 1.53 \times 10^4$. Aliquots (0.5 - 1.0 ml) were stored at -20°C. The disodium salt of CTP was purchased from P-L Biochemicals, Inc., and was prepared and stored like the ATP.

The nucleoside 5'-monophosphates pA, pC, pU and pG were purchased either from Sigma Chemical Co. or from Schwarz-Mann. The nucleoside 3'(2'), 5'-diphosphate pAp was from P-L Biochemicals, Inc.; the compounds pUp and pCp, prepared as described (108, 109), were gifts from Dr. H. G. Khorana; pGp was a gift from Dr. H. J. Gross; ribo-pTp was a gift from J. Ziegenmyer. The nucleoside 2'-O-methyl-5'-monophosphates pAm, pUm, pCm and pGm, and the modified nucleoside 3'-phosphate Ψp were purchased from P-L Biochemicals, Inc.. The modified nucleotides N^1-methyl adenosine 5'-phosphate (pm^1A) and N^7-methyl guanosine 5'-phosphate (pm^7G) were purchased from Terra Marine Bioresearch Corp. Stock solutions of nucleotides were prepared by dissolving in water and, where necessary, titrating to pH 7 with NaOH (or KOH); concentrations were generally adjusted so that the A_{260} was approximately 500. All solutions were stored at -20°C.

2', 3'-isopropylidene adenosine, used in the preparation of α [^{32}P] ATP, was purchased from P-L Biochemicals. It was dissolved in H_2O:CH_3CN = 1:2 at 8 mg/300 µl and stored at -20°C.

Glutathione, 3-(D)-phosphoglyceric acid (PGA), dithiothreitol (DTT) and spermine were from Sigma Chemical Co.. A working

stock solution of 40 mM (13 mg/ml) glutathione in water was stored at -20°C. PGA was stored at a concentration of 100 mM at -20°C; DTT was stored at a concentration of 200 mM in 20 µl aliquots at -20°C.

Ribonuclease-free A-grade bovine serum albumin (BSA) and spermidine were from Calbiochem, Inc.. BSA was dissolved in water at 1-10 mg/ml, and the concentration checked spectrophotometrically, assuming an extinction coefficient of $\mathcal{E}^{280}_{1\%} = 5.8$ (110); it was stored at -20°C. Working solutions of spermidine at 10-100 mM concentrations were stored at -20°C.

Triethylamine (TEA), purchased from Eastman Organic Chemicals, was distilled once from chlorosulfonic acid and then once over KOH (or NaOH) pellets. It was stored in amber bottles at room temperature or at 4°C. Triethylammonium bicarbonate (TEAB) was prepared by bubbling glass wool-filtered CO_2 gas through a suspension of 1150 ml of double-distilled TEA in 2.5 l water, kept near 0°C in a salt-ice bath. When the pH of the solution dropped to 7.5 - 8, the reaction was stopped and the TEAB concentration was determined by standardized titration with methyl orange indicator. The solution was adjusted to a final concentration of 2.0 M and aliquots of 0.5 - 0.8 l were stored indefinitely in opaque brown plastic bottles at -20°C or -80°C. Working solutions stored at 4°C gradually increased in pH but could be regenerated by gently bubbling through CO_2 gas for approximately 30 min..

Ultra-pure urea for polyacrylamide gel electrophoresis was from Schwartz-Mann. Acrylamide was from Eastman Organic

Chemicals and was recrystallized from chloroform before use; alternatively the Eastman "electrophoresis grade" recrystallized acrylamide was purchased. N, N'-methylene-bis-acrylamide was also from Eastman, and was recrystallized from acetone before use. Alternatively, electrophoresis grade recrystallized bis acrylamide was purchased from Bio-Rad Laboratories.

Nitrilotriacetic acid (NTA) was the "gold label" grade of Aldrich Chemical Co.. It was dissolved in water, titrated to pH 7 with NaOH and stored at 50 mM or 100 mM concentrations at -20°C or 4°C.

Crude yeast RNA for homochromatography was type VI from Sigma Chemical Co.. It was stored at 4°C.

All other chemicals were either reagent grade or the purest available commercial grade.

C. Radioactive Materials and Fluors

Carrier-free [^{32}P] orthophosphoric acid in HCl-free water, carrier-free [^{35}S] sodium sulfate in water solution, high specific activity (50-150 Ci/mmol) α [^{32}P] ATP and ultra-high specific activity (ca. 1,000 Ci/mmol) γ [^{32}P] ATP were all products of New England Nuclear Corp.. [^{14}C] UTP, at 350 mCi/mmol (New England Nuclear Corp.), was a generous gift from Dr. V. Ingram.

Two fluors were used for liquid scintillation counting of radioactive material. Toluene based scintillant contained, in toluene solution, 0.4 g/l 2, 5-diphenyloxazole (PPO) and 0.1 g/l 1, 4-bis-(2-(4-methyl-5-phenyloxazoyl)) benzene (dimethyl POPOP). Bray's

solution contained, per liter of solution in dioxane, 60 g naphthalene, 4 g PPO, 0.2 g dimethyl POPOP, 100 ml methanol and 20 ml ethylene glycol. All ingredients were either reagent grade or scintillation grade.

D. Materials for Chromatography and High Voltage Paper Electrophoresis

Whatman DEAE-paper (DE 81), purchased in rolls, Whatman 3MM, #540 and #1 paper sheets, and Whatman filter discs were from Reeve Angel and Co.. Cellulose acetate strips (3 x 57 cm) and polyethyleneimine-impregnated cellulose (PEI-cellulose) thin layer chromatography plates (glass-backed, without fluorescent indicator) were from Schleicher and Schuell, Inc.. Cellulose acetate lots were tested for minimal streaking in our usual high voltage electrophoresis conditions at pH 3.5 (see below) with the compound 5'-[^{32}P]-$(dT)_{10}$; the best lot (minimum streaking) was purchased in bulk. "Cellogel" strips were obtained through Kalex, Inc.. Plastic-backed PEI-cellulose plates (without fluorescent indicator) were Machery-Nagel (MN) brand and were obtained through Brinkmann Instruments, Inc.. Thin layer partition chromatography was performed on glass-backed 100 μ layer microcrystalline cellulose plates from E. Merck (EM).

Glass-backed DEAE-cellulose thin layer plates for homochromatography were at first made in our laboratory, but subsequently purchased from Analtech, Inc.. 20 x 20 cm plates were a 250 μ layer of cellulose MN300 HR/MN300 DEAE = 15/2, and 20 x 40 cm plates were a 250 μ layer of MN300 HR/Avicel (EM)/MN300

DEAE = 10/5/2. Cellulose powders for the home-made plates were purchased from Brinkmann Instruments, Inc.

DEAE-Sephadex A-25 and Sephadex G-50, G-75 and G-100 were supplied by Pharmacia Fine Chemicals. Whatman DEAE-cellulose DE-52 was purchased from Reeve Angel and Co.

E. RNA Samples, DNA Samples and Oligonucleotides

A sample of E. coli 5S RNA, used as a sizing marker in polyacrylamide gel electrophoresis, was a gift from Ms. S. Prakash. E. coli $tRNA_f^{Met}$ was a gift from Dr. B. Baumstark. In vivo uniformly [^{32}P]-labeled E. coli $tRNA_I^{Tyr}$ precursor, prepared according to Altman and Smith (276), was a kind gift from Dr. B. Larue. Neurospora crassa $tRNA_i^{Met}$ was a gift from Dr. A. Gillum. Euglena gracilis chloroplastic $tRNA^{Phe}$ was prepared at Oak Ridge National Laboratory, Tennessee, by Dr. L. I. Hecker (155). Brewer's yeast $tRNA^{Phe}$ was purchased from Boehringer Mannheim Corp.. Crude yeast tRNA, used as a carrier for small quantities of radiolabeled RNA or DNA, was from a stock prepared in our laboratory.

Several oligonucleotides used as sequence markers were from the library of ribotrinucleotides of Dr. H. G. Khorana (111). Other oligonucleotides were from complete T_1 or Pan RNase digests of various tRNAs from previous sequencing projects in our laboratory; sources are indicated, where appropriate, in the text (Results).

Restriction nuclease generated fragments of Ø80 pSU III DNA were generous gifts from Drs. H. Küpper and T. Sekiya and had been

prepared, originally, in the laboratory of Dr. A. Landy.

II. Instrumentation and Equipment

Cerenkov-counting of [^{32}P] was performed in a Packard model 3380 liquid scintillation counter; liquid scintillation counting was performed in the same machine or in a Beckman model LS250 counter.

Spectrophotometric determinations were made either in a Zeiss PMQ II double beam spectrophotometer or in a Gilford model 2400-S spectrophotometer. A Cary model 15 double beam scanning spectrophotometer was used for recording UV spectral analyses.

High voltage equipment for paper electrophoresis was made by Gilson Instruments, Inc.

An Eppendorf model 3200 microcentrifuge was used for centrifugation of small quantities of liquids in polyethylene or polypropylene tubes.

III. Methods

A. General Methods

1. Polyacrylamide Gel Electrophoresis and Recovery of Radioactive Material

Electrophoresis was generally performed in 10, 12 or 15% slab gels prepared with recrystallized acrylamide and N, N'-methylene-bis-acrylamide. In gels used for estimation of size of RNA, the acrylamide/bis acrylamide ratio was 40/1 (which gave a linear relationship between mobility and log of RNA chain length), while for preparative runs a ratio of 20/1 was used, which generally gave better resolution. The buffer system was the Tris-borate-EDTA

system of Peacock and Dingman (112), with the addition of urea (ultra-pure) to 7 M; no urea was used in the running buffer.

The electrophoretic apparatus generally consisted of two chambers of buffered electrolyte, with a double thickness of Whatman 3MM paper wick connecting the top of the gel to the top buffer chamber, as described by Akroyd (113). Alternatively, an apparatus allowing direct gel to buffer contact in the upper chamber, designed according to Studier (114), was used.

Samples of RNA were dissolved in 10 - 15 µl of 90 mM Tris, 90 mM borate, pH 8.3, 10 mM EDTA, 7 M urea, 25% sucrose and 0.1% each of the tracking dyes bromophenol blue and xylene cyanole FF. The solution was heated to 100°C for 2 min., chilled and then used for electrophoresis. Electrophoresis was either at room temperature or at 4°C, with a constant voltage gradient of 10 - 20 V/cm.

Radioactive bands were located by autoradiography, excised from the gel and counted (Cerenkov) in a liquid scintillation counter. The material in the excised bands was either extracted by grinding in a glass-teflon homogenizer in 0.3 M NaCl, 30 mM sodium citrate, pH 8, as published (20, 102), or was recovered by electrophoretic elution as described by Knecht and Busch (115). In the latter procedure, which reproducibly gave at least 80% recovery, electrophoresis was in pasteur pipettes with shortened tips that were plugged with 5% polyacrylamide in gel buffer. The plugged elution tubes were prerun for 3 - 12 hr at 1.5 mA/tube. Dialysis sacs containing approximately 0.5 - 1.0 ml of gel buffer were then attached, the excised gel bands were inserted, and 0.25 -

1.0 A_{260} units (12.5 - 50 ug) of carrier yeast tRNA in 25% sucrose was layered onto the surface of the plug. Electrophoresis was in fresh running buffer at room temperature for 6 - 14 hr at 1 - 2 mA/tube. The eluted radioactive material was freed of salt by extensive dialysis (2 - 3 days) against water and lyophilized; it was resuspended in small volumes of water and stored at -20°C or -80°C.

Two-dimensional gel electrophoresis was according to De Wachter and Fiers (116) with several modifications. The first dimensional polyacrylamide slab was prerun in a running buffer containing 4 - 5 mM EDTA in addition to 25 mM sodium citrate, pH 3.5; although the gel slots were filled with buffer containing 7 M urea, the tray buffer did not contain urea. For the sample run, the gel slots, too, were filled with citrate-EDTA buffer that did not contain urea. RNA samples were dissolved in 10 - 20 ul of 25 mM citrate, pH 3.5, 10 mM EDTA, 7 M urea, 25% sucrose and 0.1% each of the tracking dyes bromophenol blue and xylene cyanole FF. The samples were heated to 65°C for 5 min, chilled, and then used for electrophoresis. Electrophoresis was at 4°C with a constant voltage gradient of 10 - 20 V/cm. After autoradiography, a 15 - 16 cm section of the first dimensional gel track was excised and the second dimensional gel was polymerized directly onto it. The second dimensional buffer system also contained 5 mM EDTA in addition to 40 mM Tris-citrate, pH 8. Electrophoresis was at 4°C with a constant voltage gradient of 5 - 10 V/cm. Recovery of material from this gel was by electrophoretic elution, as described above.

2. Thin Layer Chromatography

Cellulose plates were either used directly from the box, or, for optimal resolution, were prerun in the intended chromatography solvent and dried before use. Plastic-backed PEI-cellulose plates were prerun in water and dried before use; the prerun plates could be stored up to four weeks at 4°C. Glass-backed PEI-cellulose plates, which were run in formic acid buffers, were doubly prerun before use; they were first prerun in 2M pyridinium formate, pH 2.2, dried, and then prerun in water. The prerun plates could be stored for 3 - 4 weeks at 4°C.

Samples to be spotted on cellulose thin layer plates were generally adjusted to 0.5 - 2 µl volume; DEAE-cellulose and PEI-cellulose plates could accommodate up to 10 µl sample volumes. All thin layer chromatography, except homochromatography, was run at room temperature.

The following solvent systems were prepared weekly, as needed, and stored in air-tight bottles; solvent (f), however, was freshly prepared immediately before use:

(a) isobutyric acid: concentrated NH_4OH: H_2O = 66:1:33, v/v/v (pH 3.7)

(b) isobutyric acid: concentrated NH_4OH: H_2O = 557:38:385, v/v/v (pH 4.3)

(c) 0.1 M sodium phosphate pH 6.8: ammonium sulfate: n-propanol = 100:60:2, v/w/v

(d) t-butanol: concentrated HCl: water = 70:15:15, v/v/v

(e) 0.8 M LiCl, 0.8 M acetic acid

(f) 2 M pyridinium formate, pH 3.4

(g) isopropanol:concentrated NH_4OH:0.1 M boric acid =

7:1:2 (v/v/v)

UV-absorbing material was detected under a 254 nm emitting UV light; radioactive material was detected by autoradiography. Removal of spots containing radioactive material from thin layer plates was according to the procedure of Turchinsky and Shershneva (117). The spot to be eluted was circumscribed, and a drop of solution containing 5 - 10% nitrocellulose in ethanol:acetone (1:1) was spread on it. After drying, the cellulose platelet was lifted from the plate and counted either in toluene-based scintillant, or in Bray's solution. The latter fluor was found to exhibit less quenching of [^{32}P] counts in this system; this is probably because the nitrocellulose redissolves in dioxane.

For elution, cellulose spots on thin layer plates were scraped off the surface of the plate with a flamed surgical blade and collected by suction into conical plastic Eppendorf pipette tips plugged with glass wool. Elution was by centrifugation of liquid through the packed cellulose into a tube below. Cellulose was eluted with water; PEI-cellulose was eluted with 2 M TEAB; elution of radioactive material was monitored with a geiger counter or by Cerenkov counting.

3. Fingerprinting

Fingerprinting of oligonucleotides by two-dimensional ionophoresis was according to Brownlee (99), with two minor modifications in the first dimension: 1) the cellulose acetate strip was wetted in pH 3.5 buffer that contained 2 mM EDTA in addition to 5% pyridinium acetate and 7 M urea, and 2) ionophoresis in the first dimension was run in a tank prewarmed to 30 - 35°C (by running a

small sheet of 3MM paper wetted in 5% pyridinium acetate, pH 3.5, in the tank for 15-30 min at maximum voltage, without cooling). For samples containing more than 0.2 M salt and/or more than 0.2 A_{260} units of RNA, "Cellogel" strips were used in place of cellulose acetate.

An alternative to DEAE-paper ionophoresis in the second dimension was used occasionally. It involves thin layer chromatography on polyethyleneimine cellulose using solvent (f) (118). This latter procedure is more rapid than ionophoresis on DEAE-paper, yields more compact spots and is sometimes useful for separating radioactive products which do not resolve or streak on DEAE-paper (vide infra, Results II).

Oligonucleotides were located by autoradiography; elution of material from either DEAE-paper or PEI-cellulose was in 2 M TEAB. TEAB was removed by repeated lyophilization (or repeated evaporation) from water; carrier yeast tRNA (5 - 50 µg) was usually added at this step to minimize losses of radioactive material by adsorption to plastic test tube walls.

4. Homochromatography

RNA digests ("homomixes") for DEAE-cellulose thin layer chromatography were prepared by adding KOH to 3.2 - 3.5% yeast RNA in 7.8 - 8.2 M urea (previously adjusted to pH 7.0) to a final concentration of approximately 10, 25, 50 or 75 mM KOH, and incubating at 65°C for 20 - 24 hr. After cooling to room temperature, the pH of the solution was brought down to 4.5 - 4.7 with acetic acid, and the volume then adjusted to give a final RNA concentration of 3% in

7 M urea. The less KOH in the digestion step, the greater the average chain length of the RNA in the respective solution and the stronger its elution power. Two-dimensional homochromatography was as described by Sanger and coworkers (103), using ionophoresis on cellulose acetate strips at pH 3.5 in the first dimension and DEAE-cellulose thin layer chromatography in "homomix" in the second dimension. The modifications described above for ionophoresis in the first dimension during fingerprinting were also applied in this case. Chromatography in "homomix" was at 65°C. 20 x 20 cm plates were run in glass TLC tanks and 20 x 40 cm plates were run in custom-designed plexiglass tanks built for us by Wilbur Scientific (Boston, Ma.).

5. Autoradiography

Autoradiography of paper or thin layer chromatograms, or of polyacrylamide gels used either Kodak Royal-X-O-Mat or Kodak No-Screen medical X-ray film; the latter was found two to three times more sensitive to [^{32}P] and allowed shorter exposure times. Radioactive marker ink was made by dissolving xylene cyanole FF dye in [^{35}S] sodium sulfate in water.

B. Preparation of the TYMV RNA 3'-Terminal Fragment

The following was the work of Dr. Alain Prochiantz at the University of Paris, France.

Phenoxyacetic acid-N-hydroxysuccinimide ester and [^{14}C]-Valine (225 mCi/mmol) were from Schwartz-Mann. The RPC-5 chromatography system was similar to that of Pearson et al. (119)

and consisted of polychlorotrifluoroethylene resin (Voltalef 300 LD-PL micro) coated with a quaternary ammonium salt, a trialkylmethyl-ammonium chloride (Adogen 464). The DEAE-cellulose resin (0.875 meq/mg) was from Bio-Rad, and Sephadex G-100 (coarse) from Pharmacia; benzoylated DEAE-cellulose was a gift from Dr. J. P. Waller.

Chinese cabbage leaves infected with TYMV were kindly provided by Drs. S. Astier-Manifacier and P. Cornuet. TYMV was extracted and purified according to Leberman (120), and the RNA obtained by phenol extraction.

Partially purified E. coli valyl-tRNA synthetase (Val RS) was prepared as already described (65). "RNase P" was purified from E. coli MRE 600 up to the DEAE-Sephadex chromatography step of Robertson et al. (79), using [^{14}C]-valyl-RNA of TYMV as substrate in place of $tRNA_1^{Tyr}$ precursor.

Preparation of the 3'-terminal fragment was based on the ability of the viral RNA to be aminoacylated with valine and to be cleaved by "RNase P". Purification of this fragment was based on the aminoacylation of the RNA with [^{14}C]-valine and phenoxyacetylation of the valine moiety. The purification was followed by following [^{14}C], and the phenoxyacetyl valine moiety provided the basis for the final purification step on Benzoylated DEAE cellulose. The purification scheme is summarized in Table I.

Viral RNA (70 mg) was charged with [^{14}C]-valine as already described (65). The level of acylation was about 0.37 mole per mole of RNA (assuming the molecular weight of the RNA to be 2×10^6

daltons). After incubating with "RNase P" (68), the mixture was extracted with phenol, and the RNA was recovered by ethanol precipitation and then resuspended in a minimum volume of 0.1 M NaCl, 0.01 M sodium cacodylate pH 6, before loading onto a Sephadex G-100 column (90 x 5 cm). Fractions containing the [^{14}C]-labeled RNA fragment(s)(of an average sedimentation constant of 4.5S in a sucrose density gradient) were pooled and precipitated with ethanol.

The valyl-RNA was then dissolved in water and the valyl residue phenoxyacetylated according to Gillam et al. (121) as modified by Demushkin et al. (122). The solution containing the phenoxyacetylated [^{14}C]-valyl-RNA was adjusted to the salt concentration of the starting buffer for RPC-5 chromatography.

The column (110 x 0.8 cm) was pre-equilibrated with 0.4 M NaCl, 10 mM potassium acetate pH 5, 10 mM $MgCl_2$ and 10 mM β-mercaptoethanol; chromatography was at room temperature. The RNA was applied to the column and was then eluted with a linear gradient (0.4 to 1 M NaCl; total volume: 800 ml) in the potassium acetate -- $MgCl_2$ -- β-mercaptoethanol solution. Two peaks of [^{14}C]-valine-containing material were obtained. Using the assay of Schofield et al. (123), it was found that the first peak consisted of [^{14}C]-valyl-RNA. The second peak, containing the phenoxyacetylated material, was pooled and applied to a column (5 x 0.7 cm) of benzoylated DEAE-cellulose previously equilibrated with a solution containing 1 M NaCl, 0.05 M potassium acetate pH 5 and 0.01 M $MgCl_2$. The column was washed with the same buffer and then with

Table 1. Steps in the purification of the TYMV RNA fragment.

Purification steps	RNA (mg)	cpm/mg RNA	pmol/mg RNA	Purity* (%)
Aminoacylation with [^{14}C]-valine	70	44,000	105	0.4
Cleavage of [^{14}C]-valyl-TYMV RNA with "RNase P" followed by Sephadex G-100 column chromatography	2.4	450,000	1,050	4.1
RPC-5 column chromatography	0.085	3,500,000	8,200	32
Benzoylated DEAE cellulose column chromatography	0.031	7,800,000	18,300	71

*100% purity would correspond to one pmol of valine bound per pmol of RNA of 38,000 daltons molecular weight.

2 M NaCl in the potassium acetate-$MgCl_2$ solution. A linear gradient (0 to 20%; total volume: 20 ml) of dioxane in the 2 M NaCl -- potassium acetate -- $MgCl_2$ solution was then applied. The radioactive material, which came off as a sharp peak at about 5% dioxane was pooled and recovered by ethanol precipitation. The ethanol precipitate was dried and sent to us in Boston for further analysis. Several independent TYMV RNA fragment preparations were used for the work described in this thesis.

Table I shows the steps used in the purification of the fragment and the percent yield and purity at the various stages of a typical purification experiment. Starting from 70 mg of viral RNA, a total of 31 μg of the TYMV RNA 3'-fragment was obtained after the final chromatography step on benzoylated DEAE-cellulose. Although the data in Table I, based on the ratio of [^{14}C]-valine cpm to RNA yield, indicate that the fragment is only 71% pure, more direct analyses described below (Results II) showed the fragment to be nearly homogeneous, and it was therefore used for sequence studies.

The rather low yield of the purified fragment is not entirely clear. Possible contributing factors include: 1) relatively poor aminoacylation of these RNA preparations by E. coli aminoacyl-tRNA synthetase (cf. ref. 75); 2) incomplete cleavage by preparations of "RNase P" (this phenomenon is further discussed below); 3) rather sharp cuts pooled from the various column chromatographic peaks in order to obtain more homogeneous preparations for sequence studies.

C. Preparation of $\gamma[^{32}P]$ ATP

$\gamma[^{32}P]$ ATP was prepared by minor modifications of the method of Glynn and Chappell (124). Carrier-free $[^{32}P]$ orthophosphoric acid was purchased at a concentration of 20 mCi in 50 µl of water. This was transferred to a 15 ml conical polystyrene test tube; 10 µl of concentrated buffer was added to give a final concentration of approximately 50 mM tris-HCl, pH 7.6, 7 mM $MgCl_2$, 2 mM glutathione and 0.4 mM 3-phosphoglyceric acid, and a total of 20-25 nmol of ATP. The reaction was initiated by addition of enzymes: 20 µg of glyceraldehyde-3-phosphate dehydrogenase and 10 µg of 30 phosphoglycerate kinase (in a total volume of 3 µl). Incubation was at room temperature.

The extent of exchange of $[^{32}P]$ into ATP was assayed periodically, including a zero time point (before addition of enzymes), by diluting 0.5 µl aliquots of the reaction mixture into 1.9 or 2 ml of charcoal suspension. The charcoal suspension contained 10% charcoal, 0.5 mM sodium phosphate, 0.5 mM sodium pyrophosphate and 25 mM HCl. After centrifugation, 2 µl aliquots of the supernatant were spotted onto filter discs and counted for $[^{32}P]$ in toluene-based scintillant.

The reaction was generally continued for 30 min and $[^{32}P]$ exchange yields were generally 65-80%. The reaction was stopped by dilution with water and loaded onto a small (generally 2.5 ml bed volume) DEAE-Sephadex A25 column equilibrated in 0.05 M TEAB. Unreacted inorganic $[^{32}P]$ orthophosphate was washed out in 0.1 M TEAB; a linear gradient of 0.1 - 1.0 M TEAB (total volume 100 ml)

was then applied. ATP eluted in a sharp peak at 0.4 - 0.5 M TEAB.

TEAB was removed from the pooled peak fractions by repeated evaporation from water at 35°C on a Buchler "evapomix" device. The ATP was then dissolved in water and neutralized with 4-6 µl of 0.1 N NaOH. Occasionally, a UV spectrum was taken at this point to assess the spectral purity and measure the concentration of the ATP. Radiopurity was checked by TLC on plastic-backed PEI-cellulose plates run in solvent (e). It was found that the calculated specific activity (reaction yield x total [^{32}P] added/total ATP added) agreed very well with the measured specific activity (cpm/A_{260}); in later preparations, therefore, spectrophotometric measurements were usually dispensed with to avoid unnecessary exposure to radiation. Specific activities were generally 500 Ci/mmol. The γ-[^{32}P] ATP was stored in 50% ethanol, 5 mM Tris-HCl, pH 7.5 - 8.0, at a concentration of ca. 10 µM. At -20°C, it was stable for up to 4 weeks.

D. Preparation of α-[^{32}P] ATP

The preparation of α-[^{32}P] ATP was adapted from the procedure of Symons (125), as modified by Padmanabhan and Wu (126), so as to achieve a specific activity of ca. 500 Ci/mmol. The procedure consisted of chemical phosphorylation, with [^{32}P], of 2', 3'-isopropylidene adenosine, removal of the isopropylidene protecting group, and enzymatic conversion to the triphosphate form with *E. coli* nucleotide kinase.

2', 3'-isopropylidene-adenosine was dissolved at 8 mg/300 µl

in water:acetonitrile = 1:2; carrier-free [^{32}P]-orthophosphate was diluted with non-radioactive phosphoric acid to a specific activity of 500 Ci/mmol. 15 µl of the isopropylidene adenosine solution was added to the diluted [^{32}P] and the mixture was rendered anhydrous by evaporating five times from acetonitrile. The material was then resuspended in 50 µl of vacuum-distilled dimethyl sulfoxide (DMSO) to which were added the following: 1 µl of trichloroacetonitrile (vacuum-distilled), and 1 µl of triethylamine (double-distilled; see above). Reaction was at 37°C; incorporation of [^{32}P] was assayed by diluting 1/2 µl aliquots of the reaction into charcoal suspension, as for the γ [^{32}P] ATP (see above). After 30 min at 37°C, the reaction was put on ice while the last assay time point was counted; if the reaction had gone less than 50% at this point, a second addition of trichloroacetonitrile and triethylamine was made, as before, and the reaction continued again for 15 min at 37°C. By this point, reaction yields were generally 60 -70%.

DMSO was evaporated off under high vacuum at 100°C. The residue was then resuspended in 500 µl of 5 M acetic acid and the isopropylidene protecting group was removed by heating at 100°C for 90 min in a stoppered tube. Thin layer chromatography in solvent (g) was used to confirm the quantitative removal of the isopropylidene group. Acetic acid was removed by repeated evaporation from water. Enzymatic conversion of the [^{32}P] AMP to ATP was accomplished by dissolving the deprotected nucleotide in 90 µl of 0.2 M Tris-HCl, pH 7.5, 0.04 M KCl, 0.01 M $MgCl_2$, 10 mM phosphoenolpyruvate and 5 µM non-radioactive ATP, to which

were added 30 μg of pyruvate kinase and 10 μl of crude E. coli nucleotide kinases (see above; 107). The reaction was at 37°C for 90 min; it was stored on ice while the extent of conversion of [^{32}P] AMP to ATP was checked by TLC on a PEI-cellulose plate in solvent (e). Although the reaction yield was generally greater than 90% after 90 min incubation, it was occasionally found necessary to resume incubation at 37°C (this may have been due to residual DMSO in the reaction) for several more hours until this yield was attained. Final purification of the α[^{32}P] ATP was by column chromatography on DEAE-Sephadex A25, as for γ[^{32}P] ATP (vide ante). Final storage conditions were also as for γ[^{32}P] ATP.

E. 5'-End Group Labeling of the TYMV RNA Fragment and Other Polynucleotides with [^{32}P]

The RNA or DNA was first treated with phosphatase and then labeled with [^{32}P] at its 5'-end using polynucleotide kinase.

For RNA, the incubation mixture (10 μl) for the first step contained 0.5 - 2.5 μg of RNA, bacterial or calf intestinal alkaline phosphatase (0.002 - 0.004 unit) and 50 mM Tris-HCl, pH 8.0. Incubation was at 55°C for 30 min. The phosphatase activity was then destroyed by making the solution 5 mM in nitrilotriacetic acid (NTA), incubating for 20 - 30 min at room temperature, and then for 2 - 4 min at 100°C (101). The incubation mixture (10 μl) for the second step contained 0.25 - 0.5 μg of the phosphatase-treated RNA, T4 polynucleotide kinase (2 units), 25 - 100 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$ (where necessary, additional $MgCl_2$ was added to

balance the NTA concentration), 15 mM β-mercaptoethanol and 50-100 μM γ [^{32}P] ATP. Incubation was at 37°C for 30 min. The reaction mixture was lyophilized and then subjected to polyacrylamide gel electrophoresis.

To phosphorylate intact, double-stranded DNA, the boiling step to destroy alkaline phosphatase activity was omitted. Instead, after 20-30 min at room temperature in 5 mM NTA, the solution was extracted twice with equal volumes of chloroform:isoamyl alcohol = 24:1, at room temperature. The pooled organic phases were back-extracted twice with 10 μl volumes of water, and the total pooled aqeous phases were then extracted 6 times with diethyl ether. The aqueous solution was then incubated for 20-30 min at 37°C to drive off residual ether and lyophilized. The kinase phosphorylation reaction was as above, or in a somewhat modified reaction condition described below. In general, as DNA is less susceptible to attack by unspecified nucleases contaminating various enzyme preparations, both the phosphatase and the kinase reactions could be done with as little as 1/10 the substrate used in RNA reactions, and at approximately the same enzyme levels, without appreciable reduction in yield. This was true for both native double-stranded and denatured DNA.

In some experiments, commercial phosphatase was treated with diethyl pyrocarbonate (127) before use, to reduce ribonuclease contamination. In addition, a modified polynucleotide kinase reaction condition was sometimes used in later work to obtain better yields of 5'-[^{32}P]-labeled tRNA or flush-ended DNA restriction fragments.

The enzyme (2 units) and substrate were essentially unchanged; the other components were 25 mM Tris-HCl, pH 7.6 or 8.0, 10 mM $MgCl_2$, 10 mM DTT (instead of β-mercaptoethanol), 25-75 mM γ [^{32}P] ATP, and 10% glycerol, and a final concentration of 10 γ/ml BSA. Incubation, as before, was for 30 min at 37°C.

F. 3'-End Group Labeling of TYMV RNA and of tRNA with [^{32}P]

Labeling of the 3'-end of TYMV RNA with [^{32}P] was essentially as previously described (89), except that α [^{32}P] ATP was substituted for [^{3}H] ATP. 500 µg of intact TYMV RNA was incubated in 500 µl of 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 8 mM DTT, 10 µM α [^{32}P] ATP with 30 µg of tRNA nucleotidyl transferase; the reaction was at 37°C for 90 min. TYMV RNA was recovered by ethanol precipitation and separated from unreacted α [^{32}P] ATP by gel chromatography on a column of Sephadex G-75 equilibrated in 0.02 M TEAB. TEAB was removed from the pooled void volume peak fractions by repeated evaporation from water and the RNA was stored in 500 µl of 10% ethanol at -80°C.

3'-end group labeling of tRNA was similar, except that the tRNA was first treated with snake venom phosphodiesterase (SVP) under mild conditions to remove part of the 3'-terminal C-C-A. A typical reaction used 0.1 A_{260} (5 µg) of tRNA; incubation was at room temperature in 10 µl of 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, with 0.25 µg of SVP, for 10-15 min. The reaction mixture was then extracted twice with an equal volume of phenol or with chloroform/isoamyl alcohol (24:1). After 2 back-extractions with water,

recovery of RNA was either by ethanol precipitation or by six extractions with ether, followed by lyophilization. In the subsequent reaction with tRNA nucleotidyl transferase, 0.01 - 0.05 A_{260} (0.5 - 2.5 μg) of SVP-treated tRNA was incubated in 10 μl with 1 μg of enzyme in essentially the same buffer as for TYMV RNA labeling, except that the $\alpha[^{32}P]$ ATP concentration was increased to 30 μM and 30 μM non-radioactive CTP was also added. Because of the higher concentration of nucleoside triphosphates, the reaction was faster, and incubation at 37°C was only continued for 45 min. Final recovery of 3'-end group labeled tRNA was sometimes by gel chromatography using Sephadex G-50 or G-75, but usually by gel electrophoresis.

G. Analysis for the Presence of Modified Nucleosides in RNA and DNA

This consisted of the following steps: 1) complete digestion of RNA with T_2 RNase, 2) labeling of the nucleoside 3'-phosphates thus produced with $[^{32}P]$ using polynucleotide kinase, 3) conversion of excess of $\gamma[^{32}P]$ ATP to glucose 6-phosphate with hexokinase and glucose and 4) separation of the $[^{32}P]$-nucleoside-3', 5'-diphosphates by two-dimensional TLC. In more recent work, an additional enzymatic step is included (see Results), wherein the 5'-$[^{32}P]$-nucleoside-3', 5'-diphosphates, after step 3, are converted to 5'-$[^{32}P]$-nucleoside monophosphates by treatment with nuclease P_1. These are then separated by two dimensional TLC.

For the RNase digestion, the incubation mixture (10 μl) contained 0.01 A_{260} units (0.5 μg) of RNA, T_2 RNase (0.05 units) and 10 mM ammonium acetate buffer pH 4.5. Incubation was for 5 hr at

37°C. The digest was lyophilized and redissolved in 10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 15 mM β-mercaptoethanol, 250 μM γ [^{32}P] ATP and polynucleotide kinase (2 units) in a final volume of 10 μl. The use of an alternative kinase reaction buffer, containing DTT instead of mercaptoethanol, and also BSA and glycerol (vide ante), is discussed under Results. In either buffer, the kinase reaction was generally for 30 min at 37°C. Excess γ [^{32}P] ATP was then destroyed by adding glucose to 2 mM concentration and 0.008 U of hexokinase. (This is 4 times the hexokinase concentration used in fingerprinting -- see below.) After 10 min at 37°C, 2.5 nmole of non-radioactive ATP was added; this was repeated after 10 more min, and then, after 10 final min at 37°C, the reaction was stopped by freezing.

If the [^{32}P]-labeled digest was to be further treated with nuclease P_1, it was first deproteinized by extracting twice with equal volumes of chloroform/isoamyl alcohol (as described above for tRNA). The final, lyophilized material was resuspended in 20 μl of water. 5 μl of this was incubated with 2 μg of nuclease P_1 in 75 mM ammonium acetate buffer, pH 5.3, in a total volume of 10 μl; reaction was for 3 hr at 37° C. The material was then stored frozen without further treatment.

Analysis of DNA oligonucleotides for the presence of modified nucleosides was based on the same principles as analysis of RNA, except that the initial digestion of the DNA to nucleoside-3'-monophosphates was with spleen phosphodiesterase (SP). The reaction mixture (10 μl) contained 0.01 A_{260} units of DNA, SP (5×10^{-4} U) and 50-75 mM ammonium acetate buffer; incubation was for 4-5 hr at

37°C. The solution was then deproteinized by extraction, as above, and then labeled with [^{32}P] precisely as for RNA. Further treatment with nuclease P_1 was also precisely as for RNA.

In all cases, 1 µl aliquots of the final reactions, to which were added appropriate non-radioactive carrier mononucleotides (which also served as UV markers), were used for two-dimensional separations by TLC. Analysis was on cellulose plates in solvents (a) or (d) in the first dimension, and solvent (c) in the second dimension. To maximize resolution in solvent (a), it was found useful to attach a Whatman 3MM paper wick at the top of the plate and continue development until the solvent front had run approximately 0.5 cm beyond the top of the plate (this took about 12 hr at room temperature). Additionally, it was found necessary, with either solvent (a) or solvent (d), to wait at least 24 hr, and preferably 48 hr, between the first and second dimensional runs to allow the first dimensional solvent to evaporate completely.

H. Nucleotide Sequence Determination

Conditions for the complete T_1 or pancreatic RNase digestion of the TYMV RNA 3'-fragment, subsequent treatment with bacterial alkaline phosphatase, and 5'-end labeling with γ [^{32}P] ATP and polynucleotide kinase were essentially as published (101), except that the initial RNase digestion was proportionally scaled down (18) to utilize 0.5 - 2.5 µg of RNA. The γ[^{32}P] ATP used was generally 200-500 Ci/mmol.

The incubation mixture for complete digestion (3-10 µl) contained 50 mM Tris-HCl, pH 8.0, 0.5 - 2.5 µg of RNA, and either T_1 RNase (0.2 U/µg RNA) or pancreatic RNase

(0.05 ug/ug RNA). After three hr at 37°C, E. coli alkaline phosphatase (0.5×10^{-3} U/μg RNA) was added, and the incubation was continued for two more hr at 37°C. The solution was then cooled to room temperature and 0.1 volume of 50 mM NTA was added. After 20-30 min at room temperature, the solution was heated at 100°C. for 2 min to inactivate the phosphatase, then stored frozen until needed for in vitro labeling with polynucleotide kinase.

The 5'-end group phosphorylation reaction (10 μl volume) contained 0.5 μg of RNA, treated as described above, 10 mM $MgCl_2$, 15 mM β-mercaptoethanol, 1.6 - 1.8 nmol γ[^{32}P] ATP (200-500 Ci/mmole) and 2 U T4 polynucleotide kinase. Tris-HCl buffer, pH 8.0, was not added, as it was already present in the digested RNA and in the γ[^{32}P] ATP to provide a total concentration, in the reaction, of 20-100 mM. Incubation was at 37°C. After 30 min, glucose (20 nmol) and hexokinase (0.002 - 0.003 U) were added, and incubation was continued for 10 min; 2.5 nmol of non-radioactive ATP was then added, and after 10 min more of incubation, a final 2.5 nmol of ATP was added. The reaction was incubated a final 10 min at 37°C, then stored frozen until use for two-dimensional electrophoresis. Purified oligonucleotides were phosphorylated under essentially identical conditions, except that less RNA (0.2 - 0.25 μg) was sometimes used.

5'-end group labeled oligonucleotides obtained from T_1 or pancreatic RNase digestion of RNA were "fingerprinted" as described above. Purified oligonucleotides, after 5'-end group labeling, were generally resolved from [^{32}P] glucose-6-phosphate by one-dimensional

ionophoresis on DEAE-paper in 7% formic acid. Radioactive material was located by autoradiography and eluted in 2 M TEAB; elution of DEAE paper was in a device like that described by Barrell (128), and elution of PEI-cellulose plates was as described above.

Sequence analysis of recovered oligonucleotides was in two stages, 5'-end group analysis and complete sequence analysis. For analysis of the [^{32}P]-labeled 5'-end group, 5,000 cpm of each oligonucleotide was used for complete digestion with snake venom phosphodiesterase (SVP) and an equal amount for complete digestion with T_2 RNase. The SVP reaction (4-5 µl) contained 1 µg SVP/µg carrier yeast tRNA in 50 mM Tris-HCl, 5 mM potassium phosphate, pH 8.9; incubation was for 2 hr at 37°C. The T_2 RNase reaction (5 µl) contained 0.1 U T_2/µg carrier yeast tRNA in 10-30 mM ammonium acetate buffer, pH 4.5; incubation was for 3 hr at 37°C. In either reaction, appropriate UV marker compounds were added at the end of the incubation, and the solutions were lyophilized and resuspended in 2 µl of water. Analysis of the $^{*}pN_{OH}$ end groups released by SVP and of the *pNp end groups released by T_2 RNase was by one-dimensional TLC on cellulose plates in each of two solvent systems, solvent (a) and solvent (c).

Nuclease P_1 may also be used for 5'-end group analysis of 5'-terminally labeled RNA or DNA. The reaction (10 µl) contains 1 µg P_1/1-10 µg carrier yeast tRNA in 50 mM ammonium acetate buffer, pH 5.3. Incubation is at 37°C for 5 hr. Analysis of 5'-end groups, which are identical to the 5'-mononucleotide end groups produced by SVP digestion, is by TLC, as described above.

Complete sequences of oligonucleotides were determined by partial digestion with snake venom phosphodiesterase. The enzymatic digestion was at room temperature for 160 min in 50 mM tris-HCl, 5 mM potassium phosphate, pH 8.9, with 1 µg of enzyme per 5 µg of carrier yeast tRNA; a typical reaction contained 2×10^5 cpm of [^{32}P]-labeled oligonucleotide in a total volume of 50 µl. Aliquots (5 µl) were removed at 0, 1, 2, 5, 10, 20, 40, 80 and 160 min, made 1 mM in EDTA, and heated at 100°C for 2 min to inactivate the enzyme; they were stored frozen until use. Prior to sequence analysis, the accumulation of intermediate degradation products in the aliquots was tested on small portions by one-dimensional homochromatography (see above). Aliquots were then pooled so as to best display the full range of partial degradation products. The oligonucleotides present in the partial digest were separated and identified by either one or in most cases both of the following systems:

1) one dimensional ionophoresis on DEAE-paper at pH 3.5 and/or 1.9 was used to establish the sequence of short oligonucleotides or the 5'-proximal portion of long oligonucleotides; whenever possible, partial digests of known 5'-[^{32}P]-labeled oligonucleotides were run alongside as markers;

2) sequences of longer oligonucleotides were also established by two-dimensional homochromatography (see Results).

A fuller description of these methods is given below (Results).

I. Partial Digestion of End Group Labeled RNA or DNA with Nuclease P_1

Nuclease P_1 was used to generate partial digests, both of high

molecular weight RNA and DNA, and of oligonucleotides. The incubation mixture contained 5'- or 3'-[^{32}P]-end group labeled RNA or DNA, carrier yeast tRNA and nuclease P_1 (105) at a ratio of 15 ng enzyme/ 100 µg of carrier RNA. Incubation was in 50 mM ammonium acetate buffer pH 5.3, at 20°C; native, double-stranded DNA, however, was digested at 37°C (see Results). In general, 5-10 x 10^4 cpm of material was incubated in a total volume of 50 µl using 25 µg of carrier yeast tRNA. Aliquots were removed at various times (usually 2-30 min), boiled 3-4 min in 5 mM EDTA to inactivate the enzyme, and analyzed for the extent of digestion by one-dimensional homochromatography. Appropriate aliquots were then pooled and analyzed by two-dimensional homochromatography. For high molecular weight polynucleotides, removal of aliquots at 2 min and 5 min, and occasionally also at 10 min was found most useful; for oligonucleotides, removal of aliquots at 5 min, 10 min and 20 min was generally found most useful. Where material was scarce, therefore, reactions were scaled down appropriately and only two or three time point aliquots were taken. At least 10,000 CPM was used for two-dimensional homochromatography.

RESULTS

I. Sequencing of Non-Radioactive RNA

The basic principles of sequence determination are the same for both radioactive and non-radioactive RNA. The molecule is first degraded to mononucleotides, either chemically or enzymatically, and its overall base composition is determined; for tRNA, in particular, it is also important to determine the identity and yield of all modified nucleosides. The sequencing strategy is to then degrade the molecule with base-specific ribonucleases, to separate the resulting small fragments, and to sequence these and determine their yields. Through the use of partial ribonuclease degradation of the molecule under limiting conditions, larger oligonucleotides may be isolated and analyzed; by analyzing progressively larger oligonucleotides, the small sequenced fragments may be ordered into a unique total sequence.

For determination of RNA base composition, one of the mildest degradation procedures, which does not alter most modified bases, is enzymatic digestion with the random endoribonuclease T_2. In RNA sequencing work, two base-specific ribonucleases have been used most extensively; these are T_1 RNase, which is specific for cleavage after guanine residues, and pancreatic RNase, which is specific for cleavage after pyrimidine residues. The same two nucleases are often also used for partial degradation.

If the RNA molecule to be sequenced is not radioactive, then both base composition analysis and sequence analysis must rely on spectrophotometric detection of nucleotidic material by UV absorption. To complete analysis of a tRNA sequence, for example, may therefore require over 50 mg of the purified tRNA. Also, because the large-scale fractionation of enzymatic degradation products generally uses several cycles of column chromatography, laborious chromatographic procedures may become the rate-limiting factors in a sequencing project. However, by choosing to analyze small quantitites of in vivo uniformly [^{32}P]-labeled RNA, it is possible to use more rapid fractionation procedures, such as high resolution two-dimensional paper or gel electrophoresis and two-dimensional paper or thin layer chromatography. Several such fractionation methods were pioneered by Sanger and collaborators, who also developed considerably simplified sequencing procedures (129).

Radiocative methodology was the major breakthrough that allowed rapid progress in the fied of RNA sequencing. However, many RNA species of interest, such as tRNAs from the cytoplasm or organelles of most eukaryotes or eukaryotic messenger RNA, and plant viral RNAs are difficult to label in vivo with [^{32}P] to sufficiently high specific activity for convenient analysis by these classical radiochemical sequencing techniques. To facilitate nucleotide sequence analysis of such material, our laboratory has, over the last several years, elaborated techniques for the analysis of non-radioactive

RNA phosphorylated *in vitro* with [^{32}P] by polynucleotide kinase. Using *in vitro* labeling of oligonucleotides with [^{32}P] to high specific activity, it proved possible to subsequently apply many of the same high resolution fractionation and sequencing techniques originally developed for uniformly [^{32}P]-labeled RNA (19,101). In the current work, modifications have been introduced into our established procedures and new methods have been developed so that this overall sequencing strategy is now more sensitive and considerably more efficient.

A. The Use of Polynucleotide Kinase in Modified Nucleotide Composition Analysis

The name polynucleotide kinase is a misnomer, as the enzyme, isolated from *E. coli* cells infected by bacteriophages T2 or T4, can catalyze the transfer of the γ-phosphate group of ATP to the 5'-hydroxyl group of *mononucleotides* as well as oligo- or polynucleotides (130,131). It has also been shown that the enzyme is capable of phosphorylating amino-protected deoxyoligonucleotides (132), and various modified nucleosides occurring at the 5'-ends of oligonucleotide fragments derived from tRNA. We have attempted to use T4 polynucleotide kinase, in the presence of γ[^{32}P]ATP, to phosphorylate the mixture of mononucleotides resulting from complete hydrolysis of a tRNA molecule. By subjecting the resulting 5'-[^{32}P]-labeled nucleoside diphosphates to two-dimensional TLC, we have been able to identify, in addition

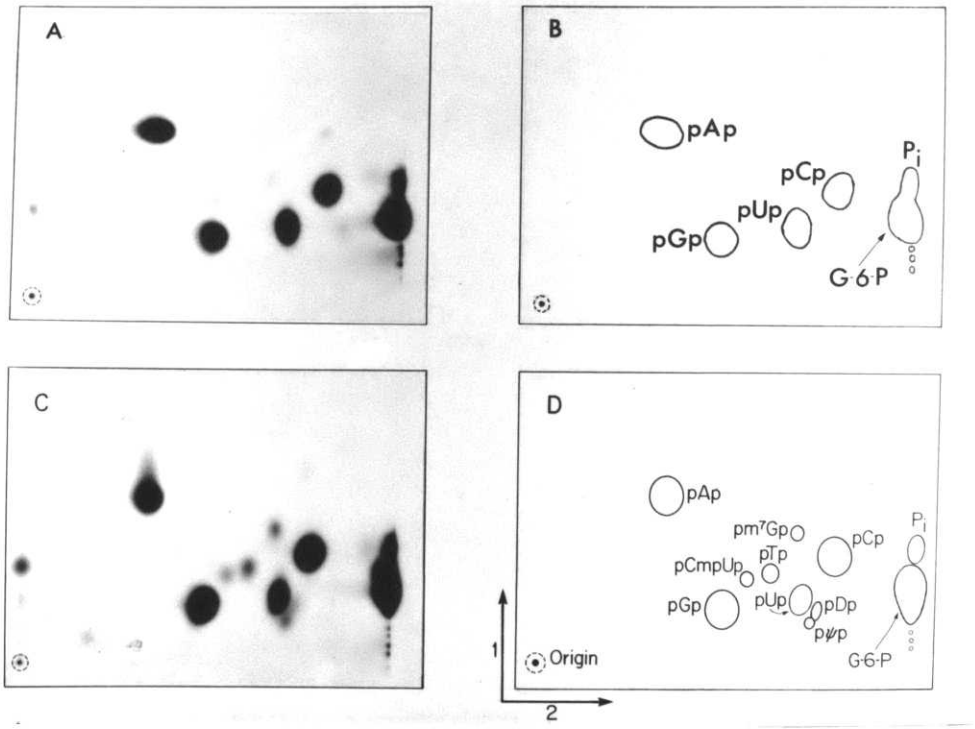
to the four major nucleotides, the modified nucleotides present in a tRNA.

For our initial work, we chose to analyze the protein synthesis initiator tRNA of E. coli, $tRNA_f^{Met}$; the unmodified tetranucleotide A-U-C-Gp served as control. Each RNA was first digested to mononucleotides with T_2 RNase. The reaction was performed in ammonium acetate buffer, which is volatile and could be removed by lyophilization afterwards. The mixtures of nucleoside 3'-monophosphates were then subjected to 5'-phosphorylation by polynucleotide kinase in the presence of $\gamma[^{32}P]ATP$. Excess $\gamma[^{32}P]ATP$ was destroyed by enzymatic transfer of the γ-phosphate group to glucose to make glucose-6-phosphate (see Methods), and the $[^{32}P]$-labeled nucleoside 3', 5'-diphosphate mixtures were analyzed by two-dimensional electrophoretic or chromatographic fractionation. Of the many two-dimensional fractionation systems that there tested, including electrophoresis on cellulose acetate followed by DEAE-paper, ion exchange thin layer chromatography on PEI-cellulose plates, and partition chromatography on cellulose thin layer plates, only partition chromatography was found to give both adequate sensitivity and adequate resolution for modified base composition analysis. In particular, it was found that complete separation of nucleotidic and non-nucleotidic $[^{32}P]$ could be achieved in solvent (c). Solvent (c) was therefore included in all two-dimensional systems tried; moreover, because solvent (c) is a high-salt, non-volatile solvent, it was always used for the second dimension. Any of several volatile sol-

vents, especially solvent (a) or solvent (d), were used for the first dimension.

The patterns obtained upon two-dimensional TLC analysis (in solvent (a) and solvent (c)) of [^{32}P]-labeled T_2 RNase digests of A-U-C-Gp and $tRNA_f^{Met}$ are shown in Fig. 2. It can be seen (Fig. 2A) that A-U-C-Gp yielded the four major nucleoside diphosphates, as expected. Moreover, as is apparent from Fig. 2C, the various modified nucleotides in $tRNA_f^{Met}$ were, in fact, phosphorylated by polynucleotide kinase. The specific identifications of the autoradiographic spots, indicated in the schematic drawings, were as follows. The four parent nucleotides were identified by comigration with the corresponding UV marker compounds. pΨp was identified by comigration with a radioactive marker compound obtained by <u>in vitro</u> phospohrylation of Ψp with polynucleotide kinase. pTp, pm^7Gp, and pCm-Up were identified by comigration with the corresponding radioactive marker compounds; these were obtained by T_2 RNase digestion (see Methods) of appropriate 5'-[^{32}P]-labeled oligonucleotides from a T_1 RNase fingerprint of <u>E. coli</u> $tRNA_f^{Met}$ (kindly supplied by Dr. B. Baumstark), i.e., from the oligonucleotides pT-ψ-C-A-A-A-U-C-C-G, pm^7G-U-C-G and pCm-U-C-A-U-A-A-C-C-C-G. All the modified nucleosides known to be present in this tRNA (133) are thus accounted for, except for s^4U, which does not appear in Fig. 2C. Its absence is probably due to loss of this labile modification (presumably by conversion to U (134))

Figure 2. Modified nucleoside composition analysis. Autoradiograms of two-dimensional TLC patterns of [^{32}P]-labeled nucleoside-3',5'-diphosphates (details in text); A) pattern obtained from the unmodified tetranucleotide A-U-C-Gp; B) replica of A), indicating identification of spots; C) pattern obtained from E. coli $tRNA_f^{Met}$; D) replica of C), indicating identification of spots. G-6-p is glucose-6-phosphate; P_i is inorganic orthophosphate. Radioactive material on the far left of plates A) and C), which does not move in the second dimension, is a mixture of radioautolysis products of γ[^{32}P]-ATP and of [^{32}P]-labeled nucleoside diphosphates; the amount of such material increases with time of storage of γ[^{32}P]ATP and/or of the [^{32}P]-labeled nucleoside diphosphate mixtures (data not shown).


A
B
pAp
pCp
pUp
pGp
Pi
G-6-P
C
D
pAp
pm7Gp
pCp
pTp
pCmpUp
pGp
pUp
pDp
pψp
Pi
G-6-P
Origin
1
2

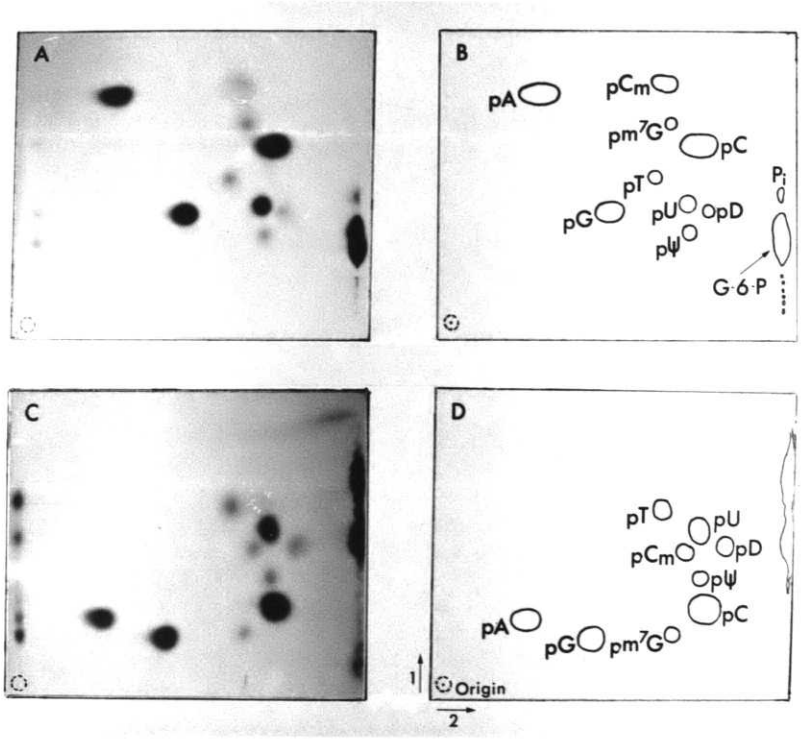
during purification and storage of the tRNA (in aqueous solution at -20°C); no s^4U could be found upon modified nucleoside composition analysis of this tRNA sample at a UV-detection scale according to Nishimura (135) (data not shown).

It may be seen, by careful inspection of Fig. 2, that both the A-U-C-Gp and $tRNA_f^{Met}$ plate display very faint additional radioactive spots not indicated in the schematic diagrams. These were identified by comigration with UV marker compounds as the four corresponding nucleoside 5'-monophosphates. It was found, in addition, that the intensity of these spots increased with time of incubation at 37°C of the T4 polynucleotide kinase reaction (data not shown); the increase in radioactivity in these spots was proportional to the loss of radioactivity from the pNp spots. We therefore assume the appearance of pN compounds to be due to the action of a 3'-phosphomonoesterase activity contaminating our preparations of polynucleotide kinase (100,136). Attempts to inhibit this activity by addition of various amounts of inorganic phosphate, spermine, spermidine or pAp into the kinase reaction mixture were not successful.

As the usefulness of our two-dimensional TLC analysis was hampered by the presence of these faint spots, an additional enzymatic step was added to our previous <u>in vitro</u> [^{32}P]-labeling procedure (see above) to <u>quantitatively</u> convert all 5'-[^{32}P] nucleoside 3',5'-diphosphates to the corresponding 5'-monophosphates. This involved treatment with nuclease P_1 (see Methods), which is known to have a 3'-

phosphomonoesterase activity (105). Enzyme was added in considerable excess (see Methods) to also insure quantitative cleavage, by the nuclease activity in P_1, of pCm-U to pCm. Results of such an experiment on $tRNA_f^{Met}$, subjected to analysis by two different two-dimensional TLC systems, are shown in Fig. 3. It is clear from Fig. 3 that all background spots have now been eliminated. Moreover, the resolution between adjacent spots is considerably improved. The identification of pCm was based on comigration with the corresponding UV marker. It can be seen that all spots in Fig. 3A except for pCm have relative mobilities very similar to those of Fig. 2C. The shift in mobility of pCm is due to its conversion from the T_2-resistant dinucleotide, pCm-Up to a mononucleotide by nuclease P_1. Parallel analysis of modified nucleoside composition by both the pNp (Fig. 2) and the pN (Fig. 3) methods can thus be used for positive identification of O-methylated nucleotides. Furthermore, as illustrated in Fig. 3C, other solvents, in this case solvent (d), can be substituted for solvent (a) in the first dimension of the two-dimensional TLC analysis. (Solvent (c) is always used in the second dimension, as discussed above, because glucose-6-phosphate and inorganic phosphate move with the solvent front.) The use of several solvent systems allows positive identification of modified nucleotides based on their mobility. The use of the pN method also allows the more general use of UV marker compounds, as these are more readily available in the pN than in the pNp form.

Figure 3. Modified nucleoside composition analysis of E. coli $tRNA_f^{Met}$, including a final treatment, prior to chromatography, with nuclease P_1 (see text). Autoradiograms of two-dimensional TLC patterns of [^{32}P]-labeled nucleoside-5'-phosphates, run in different solvent systems. A) $solvent_1$ (a), followed by $solvent_2$ (c); B) replica of A), indicating identification of spots; C) $solvent_1$ (d), followed by $solvent_2$ (c); D) replica of C), indicating identification of spots. G-6-p is glucose-6-phosphate; P_i is inorganic orthophosphate. Radioactive material on the far left of plates A) and C), which does not move in the second dimension, is a mixture of radioautolysis products of γ[^{32}P]ATP and of [^{32}P]-labeled mononucleotides (see Legend, Fig. 2).


A
B
pA
pCm
pm7G
pC
pT
pG
pU
pD
pψ
Pi
G·6·P
C
D
pT
pU
pCm
pD
pψ
pC
pA
pG
pm7G
Origin
1
2

To examine whether these methods could be extended to quantitative analysis in addition to detection, the radioactive spots in Fig. 3A were removed from the plate with nitrocellulose solution (as described in Methods) and quantitated by liquid scintillation counting.

In Table II, these results are compared with the expected molar yield of each nucleotide based on the primary sequence of this tRNA (133), and assuming quantitative conversion of s^4U to U. It can be seen that phosphorylation of modified nucleotides by polynucleotide kinase is less than quantitative. Two methods were tried to improve the yield of phosphorylation of modified nucleotides in $tRNA_f^{Met}$. The first was prolonged incubation (2 hr) at 37°C with our usual kinase reaction conditions. This gave improved 5'-phosphorylation of Ψp, Dp and Cm-U but resulted in partial conversion of m^7G to the open-ring form (137) (data not shown). Moreover, the problem of 3'-dephosphorylation, as discussed above, was considerably aggravated, so that these conditions could not be used for a two-step analysis for O-methylated nucleotides (see above); this problem also arose if higher enzyme levels were used instead of longer incubation times. The second method tried to improve enzymatic phosphorylation of modified nucleotides was to alter the reaction conditions of the kinase reaction. An optimum reaction mixture was determined empirically that substituted DTT for β-mercaptoethanol and contained, in addition, BSA and glycerol (see

Table II. Nucleoside-5'-Phosphate Molar Yields Recovered After Use of Polynucleotide Kinase in Nucleotide Composition Analysis of *E. coli* $tRNA_f^{Met}$ (Fig. 3A); A) Standard Reaction Mixture (Methods), B) Modified Reaction Mixture (Methods).

Nucleotide	Molar Yield* A	Molar Yield* B	Expected (ref. 133)
A	18	14	14
U	8	10	8+
C	24	26	25
G	23	24	24
T	1.1	1	1
Ψ	.30	.60	1
D	.16	.35	1
m^7G	.39	.46	1
Cm	.36	.52	1

*Determination by removal of spots from TLC plates (Methods) and liquid scintillation counting; in each case (A & B), results from two TLC plates are averaged. Calculation of molar yield based on:

(cpm in spot/total cpm in 9 spots) x 77

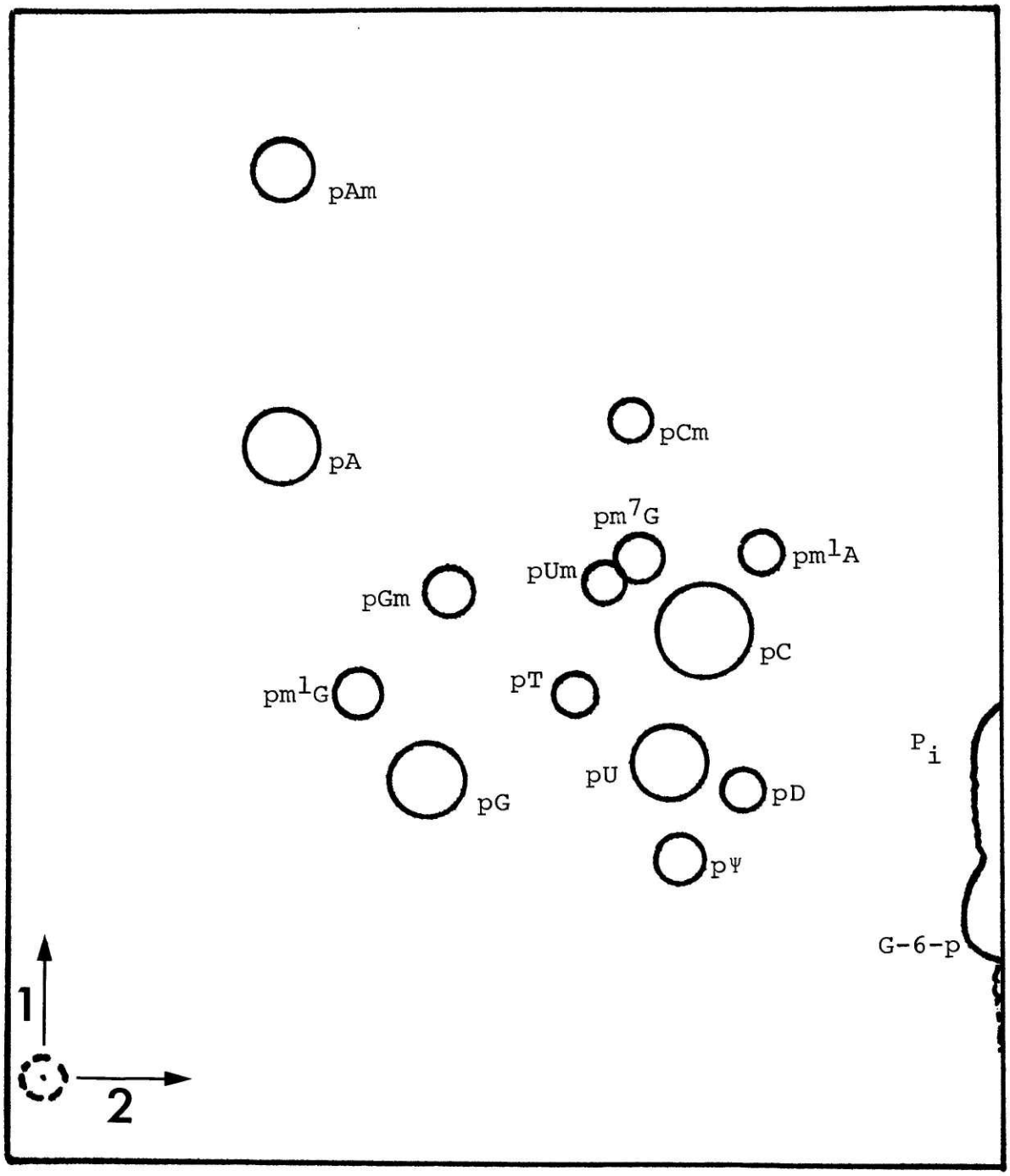
+Assuming that: a) s^4U is quantitatively converted to U (see text), and b) U in the sequence Cm-U is not counted.

Methods). Results of the application of these reaction conditions to the analysis of $tRNA_f^{Met}$ are shown in the second column of Table II. It is apparent from the data in Table II that the molar yields of [^{32}P]-labeled modified nucleotides are significantly improved. It should be mentioned in passing that spermine (2 mM), which is known to stimulate polynucleotide kinase (138), also stimulated this reaction somewhat (data not shown), and its use is still under investigation; preliminary results forecast greater than 50% phosphorylation for all the modified nucleotides in Table II.

In the course of work with other tRNAs, especially Neurospora crassa initiator $tRNA_i^{Met}$, and through the use of commercially available modified nucleoside 5'-phosphate compounds as UV markers, the positions of several additional nucleotides, not shown in Fig. 3, have been located. These are indicated in a schematic diagram (Fig. 4). It is clear that the usefulness of this analytical system can be expanded in the future by the identification of other modified nucleotide mobilities.

It was of interest to examine whether the system we had devised for RNA modified nucleoside composition analysis could also be adapted to the analysis of DNA. Such a system might be of use, for example, in the analysis of chemically synthesized oligodeoxynucleotides for purity and for extent of removal of purine and pyrimidine protecting groups (139), or for modified base composition analysis of naturally occurring DNA. The scheme that was adopted simply substituted

Figure 4. Schematic diagram of positions of modified nucleoside-5'-phosphates in the two-dimensional solvent system: $solvent_1$ (a), followed by $solvent_2$ (c). G-6-p is glucose-6-phosphate; P_i is inorganic orthophosphate.

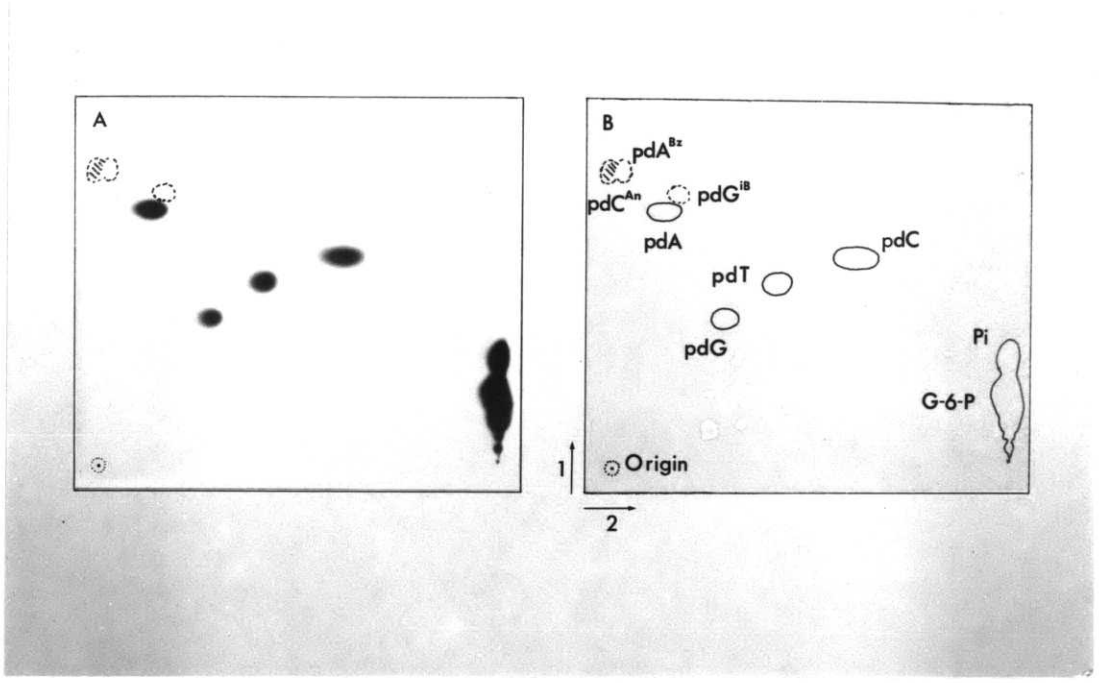

pAm
pA
pCm
pm7G
pUm
pm1A
pGm
pC
pm1G
pT
Pi
pU
pD
pG
pΨ
G-6-p
1
2

spleen phosphodiesterase for T_2 RNase in the initial digestion of the polymer to mononucleotides, as described in Methods. An analysis of dG-C-A-T-C-A-T-A-T-C-A-A-A (kindly provided by Dr. B. Ramamoorthy) is shown in Fig. 5; the mononucleotides were identified by comigration with corresponding UV marker compounds. The mobilities of nucleotides containing amino-blocking groups typically used in organic chemical synthesis of DNA (139) are also indicated in Fig. 5. While these results indicate that these procedures can be applied to qualitative analysis of DNA nucleotide composition, quantitative analysis of the radioactive material in the TLC spots showed greater deviations from expected results than was found for RNA (data not shown); the cause of this phenomenon is not known. DNA analysis at the pNp level, i.e. without the nuclease P_1 step, was not possible because of considerable partial conversion to pN during the kinase reaction (data not shown); this may be due to the preference of the implied 3'-nucleotidase for deoxy over ribonucleotides (136).

B. The Use of Polynucleotide Kinase in Fingerprinting

Sequence analysis of RNA generally begins with T_1 RNase and pancreatic RNase digestions of the substrate, and fractionation of the resulting oligonucleotides. For non-radioactive RNA, an additional step of radioactive post-labeling of these oligonucleotides is introduced. The procedures used in our laboratory for the introduction of [^{32}P] at the 5'-ends of non-radioactive oligonucleotides (18,101) are modifications of

Figure 5. Modified nucleoside composition analysis of the oligodeoxynucleotide dG-C-A-T-C-A-T-A-T-C-A-A-A; TLC in solvent$_1$ (a) followed by solvent$_2$ (c). A) Autoradiogram; B) replica of A), indicating identification of spots. The UV marker compounds pdA^{Bz}, pdC^{An} and pdG^{iB} are also indicated. G-6-p is glucose-6-phosphate; P_i is inorganic orthophosphate. Shading indicates fluorescence under exposure to 254 nm UV light.


A
B
pdA^Bz
pdC^An
pdG^iB
pdA
pdC
pdT
pdG
Pi
G-6-P
Origin
1
2

methods originally described by Szekely and Sanger (100). The sequence of operations is: 1) T_1 or pancreatic RNase digestion, 2) concurrent removal of 3'-phosphomonoester groups, in the oligonucleotides produced, with alkaline phosphatase, 3) inactivation of phosphatase by heating in the presence of a chelator, NTA, 4) introduction of a 5'-phosphate group by treatment with T4 polynucleotide kinase in the presence of $\gamma[^{32}P]$ATP and 5) quantitative elimination of excess $\gamma[^{32}P]$ATP by phosphorylation of glucose to $[^{32}P]$-glucose-6-phosphate with yeast hexokinase (a procedure first described, in this context, by Wu (140)). Experimental details are given in Methods.

The material that results from the sequential operations described is a mixture of oligonucleotides bearing 3'-hydroxyl groups and an external $[^{32}P]$-5'-phosphomonoester group. Mononucleotides present in the original nuclease digestion are mostly converted to nucleosides by alkaline phosphatase, and therefore cannot act as substrates for enzymatic phosphorylation by polynucleotide kinase. However, the kinetics of the RNase digestions is such that some mononucleoside-2',3'-cyclic phosphate intermediates, which are resistant to alkaline phosphatase, remain at the end of the incubation. These are subsequently phosphorylated by polynucleotide kinase, and are therefore also present among the $[^{32}P]$-labeled oligonucleotides. The cyclic phosphate group eventually opens, possibly due to the continued presence of RNase throughout

all steps or to the work-up procedures, and the mononucleotides are recovered as *pNp (*p = [^{32}P]).

The mixture of 5'-labeled oligonucleotides resulting from T_1 RNase or pancreatic RNase digestions of an RNA molecule may be fractionated by the two-dimensional electrophoresis techniques (see Methods) originally developed by Sanger and collaborators for similar analyses of uniformly [^{32}P]-labeled RNA (129). Because the mobilities of oligonucleotides in these systems are very reproducible and highly sensitive to minor variations in nucleotide composition or sequence, the fractionation pattern resulting from base-specific nuclease analysis of an RNA molecule has been termed a "fingerprint". Comparison of T_1 RNase and/or pancreatic RNase fingerprints of different RNA molecules is a very sensitive technique for studying sequence relatedness. Application of the polynucleotide kinase [^{32}P]-post-labeling procedure allows even very limited quantities of non-radioactive RNA to be studied by fingerprinting; moreover, as the 5'-[^{32}P]-labeled oligonucleotides fractionated by fingerprinting can be recovered and sequenced, this also provides a means to the first step of sequence analysis of non-radioactive RNA. Through the use of γ[^{32}P]ATP of high specific activity (200-500 Ci/mmol), and by applying the efficient sequencing procedures discussed below, the sequences of all the oligonucleotides resulting from its T_1 RNase or pancreatic RNase digestion can be derived from as little as 10-15 pmol of an RNA molecule

(for a tRNA molecule, this corresponds to ca. 0.01 A_{260} units, or 0.5 μg).

Examples of the use of 5'-terminal labeling in fingerprinting are the T_1 and pancreatic RNase fingerprints of the TYMV RNA 3'-fragment, shown in Fig. 16 (below), and of Euglena gracilis chloroplastic $tRNA^{Phe}$, shown in Fig. 29, and the pancreatic RNase fingerprints of E. coli 5S RNA, shown in Fig. 33, below.

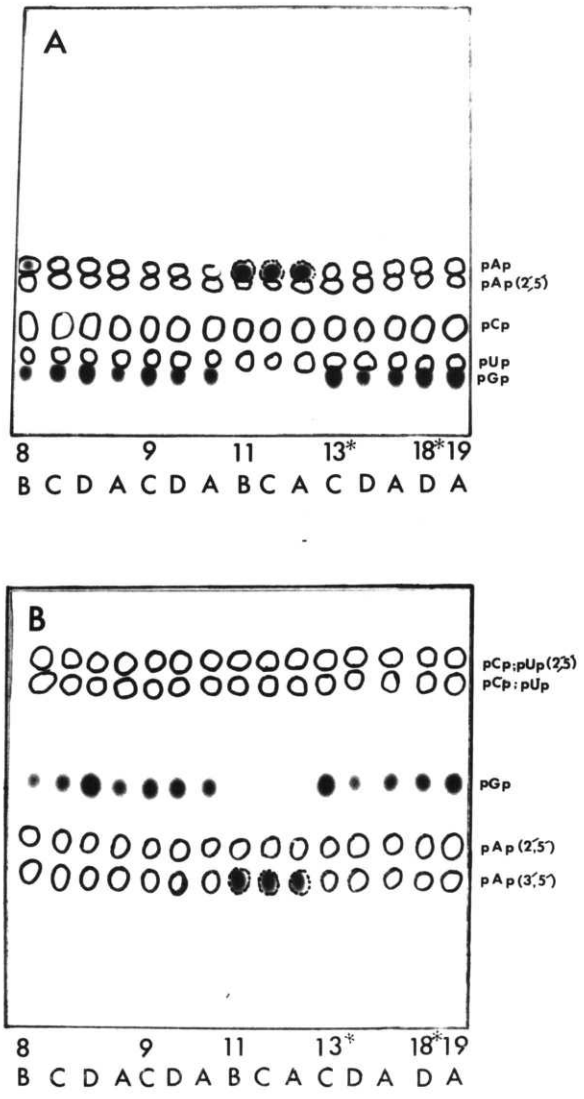
C. Sequence Analysis of 5'-[^{32}P]-labeled Oligonucleotides

For each oligonucleotide recovered from a fingerprint, the identity of the 5'-end nucleotide, carrying the [^{32}P]-label, can be determined by complete enzymatic degradation to mononucleotides, followed by thin layer chromatography and autoradiography, as described in Methods. To confirm each 5'-end group analysis, we generally perform independent digestions with two enzymes, RNase T_2 and snake venom phosphodiesterase, and analyze each digest in two solvent systems, (solvent (a) and solvent (c)). RNase T_2 releases a [^{32}P]-labeled 5'-end mononucleoside 3',5'-diphosphate of the form *pNp, while snake venom phosphodiesterase releases a [^{32}P]-labeled nucleoside 5'-monophosphate of the form $*pN_{OH}$. Location of radioactive spots is by autoradiography, and identification of mononucleotides is by comigration with corresponding UV marker compounds. An example of a typical

5'-end group analysis of fingerprint oligonucleotides is illustrated in Fig. 6. Wherever possible, 5'-end group modified nucleotides, too, are identified with UV marker compounds; however, dihydrouridine, which has no UV absorbance, and several other common and less common modified nucleotides, for which no markers are on hand, must be identified by R_f alone. The use of chromatography in several solvent systems, and also additional analysis by ionophoresis on paper or DEAE-paper at pH 3.5, are occasionally useful in confirming such identifications.

The sequence of the nucleotide residues succeeding (5'→3') the 5'-end residue of an oligonucleotide is determined by partial digestion of the oligonucleotide with snake venom phosphodiesterase (SVP). This enzyme is an exonuclease which recognizes the 3'-hydroxyl end of an oligonucleotide and successively removes nucleoside-5'-phosphate residues from the 3'-end (141). When digestion conditions are mild and well-controlled, accumulation of partial degradation products is gradual. By stopping aliquots of the reaction at appropriate timed intervals, a range of 5'-[^{32}P]-labeled oligonucleotides is obtained, covering every intermediate from the 5'-end group to the starting material. Partial digestion of fingerprint oligonucleotides with SVP is performed as described in Methods, and the accumulation of 5'-[^{32}P]-labeled partial degradation products is assayed by one-dimensional homochromatography. An example of such a

Figure 6. Example of 5'-end group analysis following digestion of 5'-[^{32}P]-labeled oligonucleotides with T_2 RNase. Autoradiograms of one-dimensional TLC analyses in A) solvent (a), and B) solvent (c); the circles indicate positions of UV marker compounds (identified alongside the pattern) run together with the radioactive samples. The numbered TLC tracks correspond to the numbered spots in the pancreatic RNase fingerprints of 5S RNA in Fig. 33 A-D, and to the numbered oligonucleotides in Table VIII.

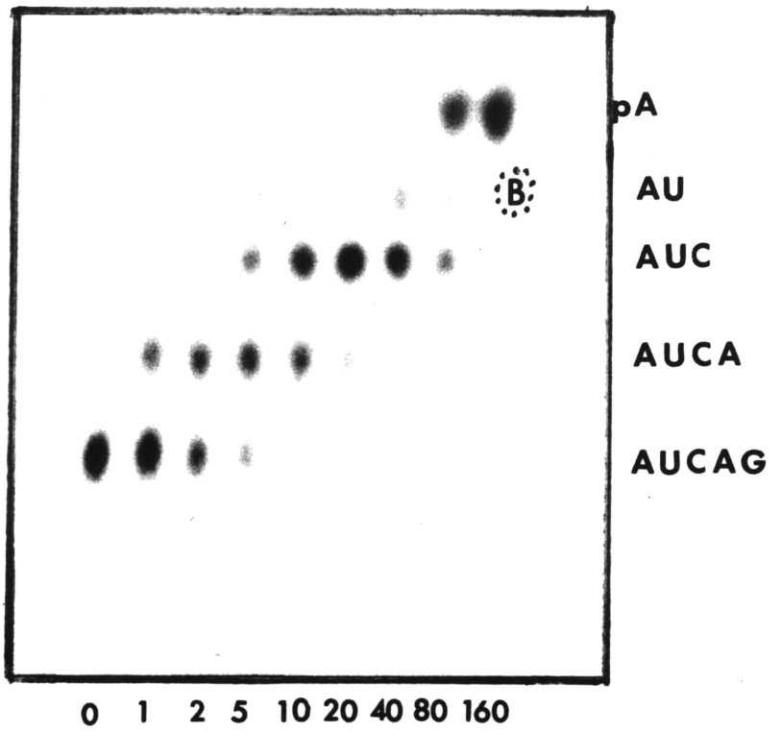

A
pAp
pAp(2′,5′)
pCp
pUp
pGp
8 9 11 13* 18*19
B C D A C D A B C A C D A D A
B
pCp;pUp(2′3′)
pCp:pUp
pGp
pAp(2′,5′)
pAp(3′,5′)
8 9 11 13* 18*19
B C D A C D A B C A C D A D A

partial digest is illustrated in Fig. 7; the oligonucleotide here analyzed corresponds to the A-U-C-A-G marker compound of Fig. 9A,B, below. The standard digestion conditions we have used, which include a wide range of time points, have generally been found sufficient for almost all oligonucleotides from two to 13 nucleotides in length; exceptional cases, involving 3'-end or internal modified bases, are discussed below.

Sequence information about an oligonucleotide may be derived from analysis of the mobilities of its homologous series of terminally labeled partial degradation products in any of several electrophoretic and chromatographic systems. We have used two systems in particular, ionophoresis on DEAE-paper at pH 3.5 and/or 1.9, and two-dimensional homochromatography.

Ionophoretic analysis, on DEAE-paper, of an SVP-generated series of oligonucleotide partial degradation products is based on use of the pH 3.5 and pH 1.9 buffer systems described by Sanger and collaborators (99,129). At either pH, the ionophoretic mobility of the shorter oligonucleotides of a homologous series is greater than that of the larger. Mobility shifts between immediate homologues are characteristic of the specific nucleotide by which they differ. These mobility shifts can be quantitated as "M-values", wherein the distance from the origin of two successive homologues, X and Y, is used to calculate the characteristic value $M = \frac{Y-X}{X}$

Figure 7. Analysis of the progress of the partial digestion, with SVP, of 5'-[^{32}P]-labeled *pA-U-C-A-G. Autoradiogram of one-dimensional homochromatography analysis, indicating identification of compounds; time in min. B (surrounded by dotted circle) is xylene cyanole dye. Note the smaller pyrimidine mobility shift, pA-U-C $\longrightarrow$ pA-U, as compared with the distance of the purine mobility shifts pA-U-C-A $\longrightarrow$ pA-U-C, and pA-U-C-A-G $\longrightarrow$ pA-U-C-A.


pA
B
AU
AUC
AUCA
AUCAG
0 1 2 5 10 20 40 80 160

(99,129). Table III reproduces the empirical M-value ranges for the four major nucleotides which were published by Sanger and collaborators; deviations from these M-value ranges were observed, occasionally, in our own work with 5'-terminally labeled oligonucleotides, and are indicated in parentheses.

Only a limited size range of oligonucleotides will give M-values within the ranges indicated in Table III. Shorter oligonucleotides, which have mobilities beyond that of xylene cyanole, the tracking dye, do not give proper M-values, for example, although their mobilities are still very characteristic and reproducible. Such short oligonucleotides are generally identified by comparison of their mobilities with those of oligonucleotide standards of known sequence (often run alongside the unknown); 5'-end group mononucleotides are identified with the corresponding UV marker compounds. Oligonucleotides longer than 5-7 nucleotides are usually so crowded near the origin that M-values cannot be accurately determined. Positive sequence analysis by this method is therefore limited to the size range 1-6 nucleotides.

Analysis of the sequence of oligonucleotides recovered from fingerprints is simplified by the fact that they result from base-specific cleavages; T_1 RNase fingerprint oligonucleotides have only one G residue, which is at the 3'-terminus, and pancreatic RNase fingerprint oligonucleotides have only one pyrimidine residue, which is at the 3'-terminus. Nevertheless, to eliminate occasional ambiguities inherent in the

Table III. Ranges of M Values on DEAE-Paper Ionophoresis at pH 3.5 or 1.9 (99,129).

3'-Terminal Nucleotide Removed	Value of M	
	pH 3.5	pH 1.9
pC	0.6-1.2	0.05-0.3
pA	2.1-2.9 (4.1)*	0.4-1.1
pU	1.7-1.9 (2.5)*	1.5-2.5
pG	2.6-4.4	1.2-3.1

*Numbers in parentheses indicate deviations from the M-value ranges published by Sanger and collaborators (99,129) which were observed, occasionally, in our own work with 5'-terminally labeled oligonucleotides.

overlap or near-overlap of the M-values of A, U and G at pH 3.5 and of U and G at pH 1.9 (Table III), most sequence identifications are confirmed at both pH 3.5 and pH 1.9. In addition, oligonucleotide standards of known sequence are run alongside the unknown whenever possible; this last refinement is also essential in identifying mobility shifts due to modified nucleotides.

These points are illustrated in the sequence analysis of several T_1 fingerprint oligonucleotides in Figs. 8 and 9. Fig. 8 shows the analysis of 5'-terminally labeled pU-U-C-U-C-G alongside the known oligonucleotide pU-C-A-U-C-G (corresponding to spots 15 and 12 in the T_1 RNase fingerprint of Fig. 16A, below) at pH 3.5 (Fig. 8A) and pH 1.9 (Fig. 8B). It can be seen that the 3'-proximal C mobility shift (pU-U-C-U-C $\longrightarrow$ pU-U-C-U) at pH 1.9 (Fig. 8B, left) is so small that this spot does not resolve from the next shorter oligonucleotide in the series. Positive sequence identification, in this case, required analysis at pH 3.5 (Fig. 8A). Conversely, in the example illustrated in Fig. 9, the data at pH 3.5 (Fig. 9A) were ambiguous, and required clarification by analysis at pH 1.9 (Fig. 9B). The oligonucleotide in question here is pA-U-C-U-U-U-A-A-A-A-U-C-G, which corresponds to spot 16 in the T_1 RNase fingerprint of Fig. 16A, below; it is run alongside a standard of known sequence, pA-U-C-A-G (cf. Fig. 7). It can be seen (Fig. 9) that the mobility shifts between A-U-C and A-U-C-A or A-U-C-U are essentially identical at pH 3.5 (Fig. 9A), yet are clearly distinguished

Figure 8. Autoradiograms of sequence analysis by one-dimensional DEAE-paper ionophoresis at A) pH 3.5 and B) pH 1.9 of the 5'-terminally labeled oligonucleotide i) *pU-U-C-U-C-G run alongside the known oligonucleotide ii) *pU-C-A-U-C-G (see text for details). The tracks labeled o are examples of each oligonucleotide before addition of enzyme, and the successively numbered tracks are pooled time point aliquots of a partial SVP digestion of each oligonucleotide (c.f. Fig. 7) selected so as best to represent the progress of the reaction.

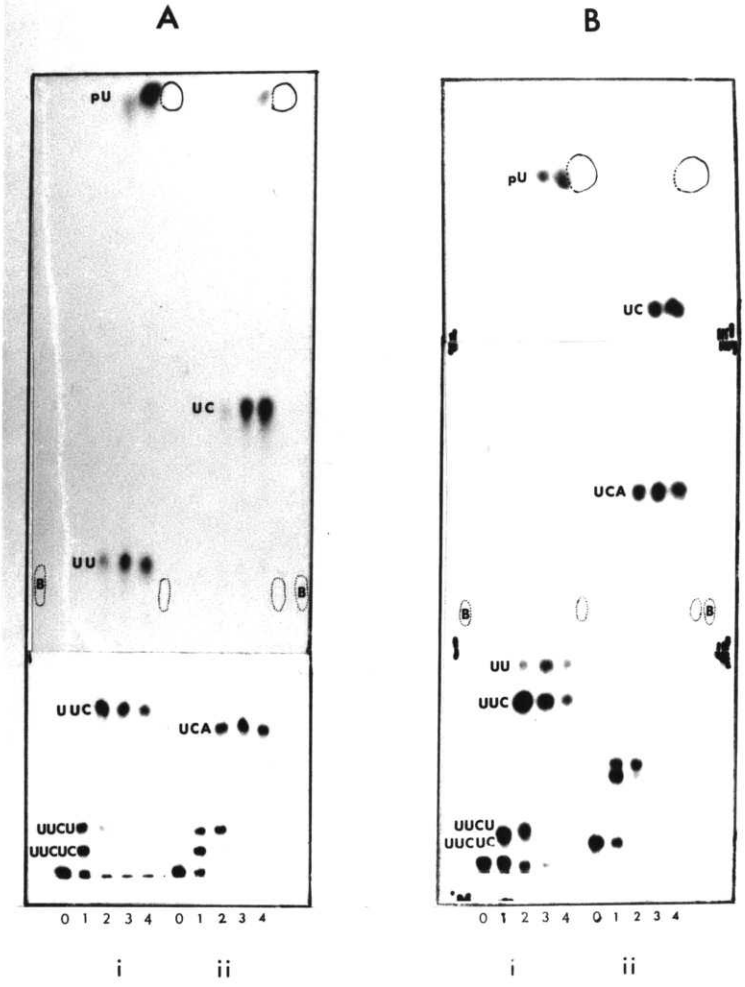
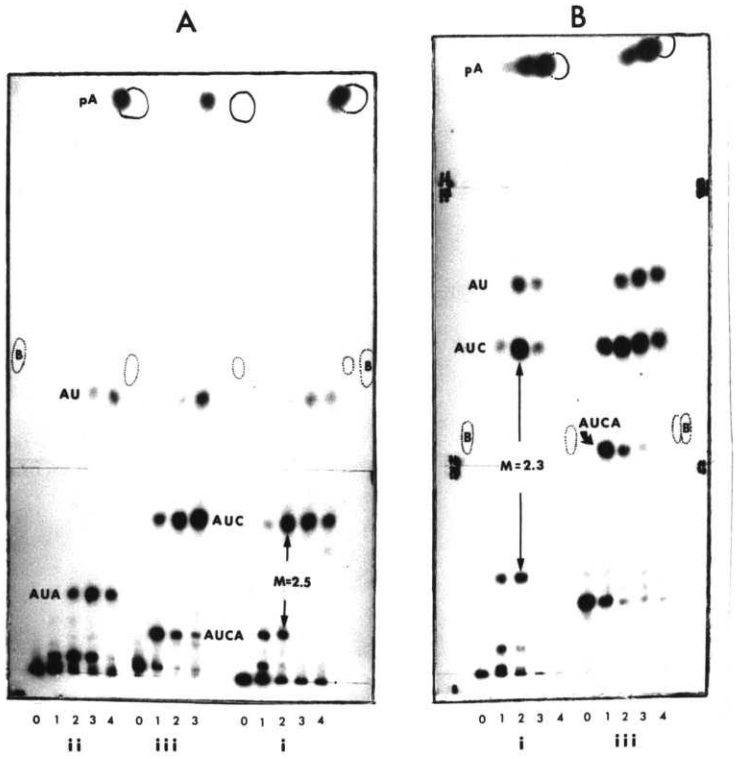

A
B
pU
UC
UU
UUC
UCA
UUCU
UUCUC
UCA
0 1 2 3 4 0 1 2 3 4
i ii
pU
UC
UCA
UU
UUC
UUCU
UUCUC
0 1 2 3 4 0 1 2 3 4
i ii

Figure 9. Autoradiograms of sequence analysis by one-dimensional DEAE-paper ionophoresis at A) pH 3.5 and B) pH 1.9 of the 5'-terminally labeled oligonucleotide i) *pA-U-C-U-U-U-A-A-A-A-U-C-G run alongside the known oligonucleotides ii) *pA-U-A-A-U-C-G and iii) *pA-U-C-A-G (see text for details). The numbered tracks represent the progress of a partial digestion of each oligonucleotide with SVP (cf. legend to Fig. 8). Note that the M-values of interest (see text) are indicated.

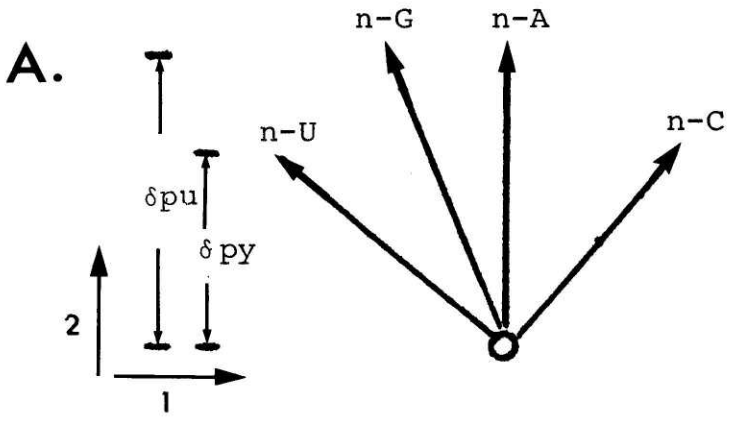

A
pA
AU
AUC
AUA
AUCA
M=2.5
0 1 2 3 4 0 1 2 3 0 1 2 3 4
ii iii i
B
pA
AU
AUC
AUCA
M=2.3
0 1 2 3 4 0 1 2 3 4
i iii

pH 1.9 (Fig. 9B). Fig. 9 also illustrates that, even for oligonucleotides too long to sequence completely by ionophoresis, very useful data about 5'-proximal sequences can nonetheless be obtained. In most of the sequence work described below, ionophoresis of partial SVP-digests of long oligonucleotides was routinely used, in this manner, to analyze 5'-proximal sequences.

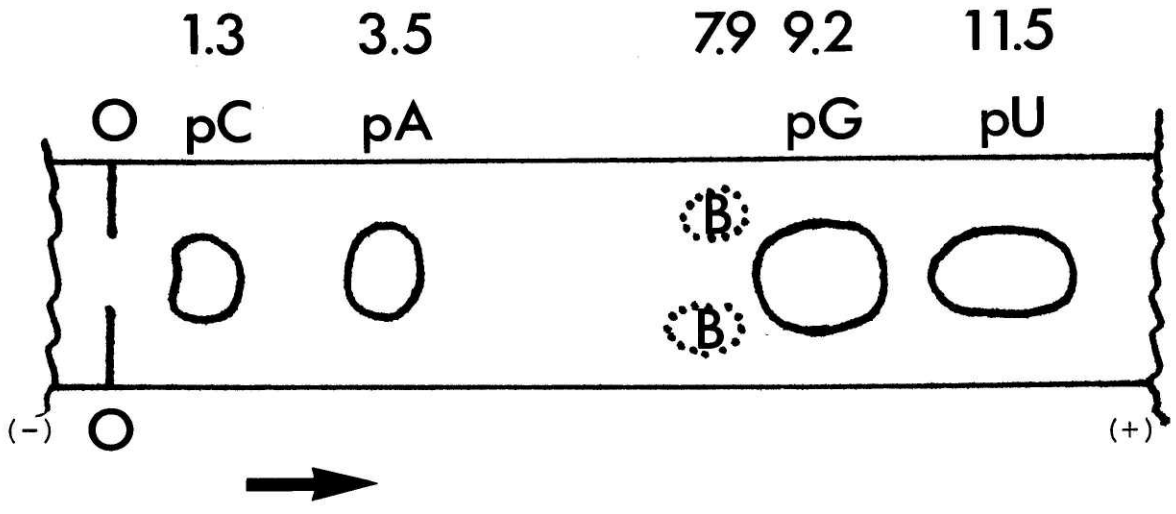
The second method we have used for sequence analysis of homologous series of partial oligonucleotide degradation products is two-dimensional homochromatography. This involves ionophoresis on cellulose acetate at pH 3.5 in the first dimension, and DEAE-cellulose TLC in homomix in the second dimension (details in Methods). The use of this two-dimensional fractionation system as a sequencing method is based on the DNA sequencing work of Ling (142), Ziff et al. (143) and Sanger et al. (103) and on the RNA sequencing work of Rensing and Schoenmakers (144). It is a method particularly useful for sequencing longer oligonucleotides (see below).

In the first dimension, ionophoresis on cellulose acetate at pH 3.5, relative mobilities of oligonucleotides within a homologous partial degradation series are characteristic of the nucleotide by which they differ. The removal of pU, pG, pA or pC from an oligonucleotide results in a characteristic increased or decreased ionophoretic mobility (Fig. 10A), the effect of each nucleotide being a function of its pKa (145,146). The pKa values of the four major

Figure 10. A) Direction of angular mobility shift in two-dimensional homochromatography upon removal of one nucleotide from an oligonucleotide. B) Mobilities of the four nucleoside-5'-phosphates in ionophoresis on cellulose acetate at pH 3.5; detection was by UV absorbance. Ionophoresis was for 30 min at 5 KV and ca. 30°C. B (circled) is xylene cyanole dye. Distances of migration are in cm., beginning at O (origin).



B.



nucleotides, and the calculated average charge of each nucleotide at pH 3.5 (147) are indicated in Table IV. Also listed in Table IV (last column) are the experimentally observed relative mobilities of the four major nucleotides on cellulose acetate at pH 3.5 and ca. 30°C (Fig. 10B); it may be inferred from a comparison of calculated nucleotide charge and experimentally observed mobility that the ionophoretic mobilities of mononucleotides on cellulose acetate at pH 3.5 are based mostly on their charge. For larger molecules, such as oligonucleotides, there is also a "drag" effect related to size and shape of the molecule, and to solvent counterion effects (145, 146). The observed effects of the four major nucleotides on oligonucleotide electrophoretic[1] mobility may be explained as the combined contributions of charge and mass (i.e., "drag"). For example, at pH 3.5, pC is only very slightly negatively charged (Table IV). In an electric field, therefore, this residue will, in most cases be causing drag on the oligonucleotide by adding mass without enough charge to compensate; however, for very slightly charged compounds, the increase in charge may still dominate. As a practical example, pU-C will run considerably more slowly than pU, whereas pC-C will run slightly faster than pC (in this latter case, note that the increase in charge is two-fold). The nucleotides pU and pG have a full negative charge at pH 3.5, and their addition to an oligonucleotide substantially increases electrophoretic mobility (or vice versa for their removal (Fig. 10A)). The

[1]The term electrophoretic is preferred over ionophoretic for large molecules.

Table IV. Calculated and Experimentally Determined Relative Mobilities (Relative Charges) of 5'-Nucleotides at pH 3.5.

Nucleotide	pKa[a]	q (calc.)[b]	q (obs.)[c]
pU	--	-1	
pG	2.4	-0.92	-0.80
pA	3.74	-0.34	-0.30
pC	4.5	-0.10	-0.11

[a]Ref. 147.

[b]q (calc.) = the charge on the compound at pH 3.5 calculated from the relation (and assuming the contribution of the 5'-phosphate group to be -1):

$$pH = pKa + \log \frac{[base]}{[acid]}$$

[c]q (obs.) = the charge on each compound obtained from the data of Fig. 10B by assigning the mobility of pU a value of -1 and assuming that the relative mobility of each compound is proportional to its relative charge. The data of Fig. 10B were obtained at pH 3.5, ca. 30°C (see Legend, Fig. 10B).

effect of pA is between the two extremes; its partial net negative charge essentially compensates for the mass it adds to an oligonucleotide chain, resulting in little change in electrophoretic mobility (Fig. 10A).

The second dimension of this two-dimensional fractionation system, DEAE-cellulose thin layer chromatography in "homomix" (see Methods), resolves an oligonucleotide mixture on the basis of size, shorter oligonucleotides moving faster than longer oligonucleotides. In addition, as can be seen in Fig. 17 (and see legend there), a difference of one purine residue between two immediate homologues in a series of partial degradation products causes a larger mobility shift than a difference of one pyrimidine residue. The overall result in two dimensions is that the stepwise removal of nucleotides in a homologous partial degradation series can be "read" as anglular mobility shifts between successively shorter fragments. The identity of a similar first-dimensional U or G mobility shift, for example, is confirmed by the pyrimidine or purine character of the second-dimensional component of the mobility shift (and likewise for C and A shifts).

By very careful standardization of experimental conditions Wu and colleagues have succeeded in controlling the electrophoretic mobilities of oligonucleotides on cellulose acetate to the point of mathematical predictability (145,146); in their system, calculated oligonucleotide mobilities, based on multi-

plication of the total theoretical charge on an oligonucleotide by an empirically derived size factor, are used to confirm mobility shift identifications. Because our high voltage electrophoresis equipment (see Materials) is not amenable to the requisite rigorous temperature standardization, we have not attempted mathematical sequence analysis. However, by accumulating considerable reference data on the behavior of homologous partial degradation series of oligonucleotides of known sequence, including oligonucleotides that contain modified bases, we have been able to use two-dimensional homochromatography as an entirely self-sufficient and unambiguous method of sequence analysis. It may be noted in passing that many of the sequence data on oligonucleotides derived from the TYMV RNA 3'-fragment, discussed below, were obtained before this two-dimensional homochromatography reference data was available; the 5'-proximal portions, at least, of all these sequences were therefore confirmed by ionophoretic analysis, as described above. Today, on the other hand, these TYMV sequences have themselves become a major addition to our library of homochromatography reference data.

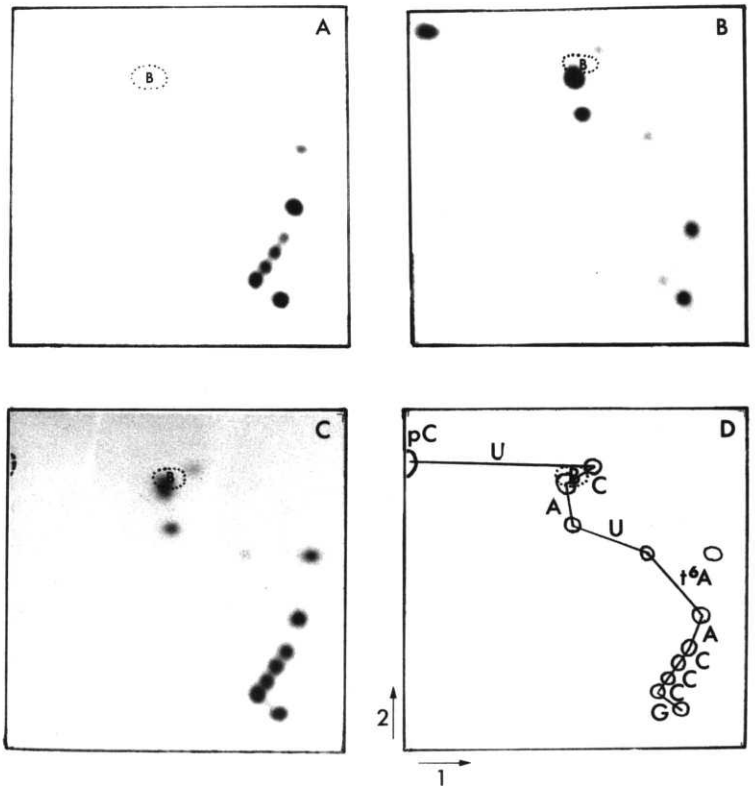
D. Partial Digestion of 5'-[^{32}P]-labeled Oligonucleotides with Nuclease P_1

Most oligonucleotide sequences can be determined by partial enzymatic degradation with SVP. A special problem arises, however, with oligonucleotide sequences that contain 3' or internal modified nucleoside residues that block the stepwise

3'→ 5' exonucleolytic progress of SVP. In such cases, the terminally-[^{32}P]-labeled partial degradation products resulting from cleavage at phosphodiester bonds on the 5' side of the modified nucleoside can be obtained in better yield by partial endonucleolytic digestion with nuclease P_1. Since both SVP and nuclease P_1 cleave phosphodiester bonds to leave 3'-hydroxyl and 5'-phosphate ends, aliquots of partial digests produced by each of these enzymes (see Mehtods) can subsequently be combined to give an optimal representation of all the terminally-[^{32}P]-labeled partial degradation products of an oligonucleotide. The combined partial digests are analyzed by two-dimensional homochromatography.

The use of this method is illustrated in Fig. 11, which represents the nucleotide sequence anlaysis of the T_1 oligonucleotide resulting from the anticodon region of Neurospora crassa $tRNA_i^{Met}$ (20), pC-U-C-A-U-t^6A-A-C-C-C-G. The SVP digest (Fig. 11A) is blocked at the t^6A residue. The nuclease P_1 digest, on the other hand, being an essentially random endonucleolytic digest, accumulates all intermediates except the one resulting from cleavage at the t^6A-A phosphodiester bond (Fig. 11B). The two digests are therefore complementary, and when pooled and analyzed together yield the complete sequence of the oligonucleotide (Fig. 11C). The nature of the t^6A mobility shift, and the identification of one artifactual spot (Fig. 11) was based on previous work of A. Gillum (20).

Figure 11. Analyses by two-dimensional homochromatography of partial digests of an oligonucleotide generated by T_1 RNase from the anticodon region of Neurospora crassa $tRNA_f^{Met}$: *pC-U-C-A-U-t^6A-A-C-C-C-G (20). Autoradiograms of patterns obtained from A) a partial digestion with SVP, B) a partial digestion with nuclease P_1, and C) combined aliquots of the digests indicated in A) and B). D) is a replica of C), indicating identification of mobility shifts. Note the spot near the t^6A shift, which is artifactual, and is generated by an endonuclease contaminating SVP, which leaves 3'-phosphate end groups; although the sequence of this material has not been determined, its artifactual nature has been confirmed by comparison with similar spots appearing in the homochromatographic sequence analysis of the known anticodon oligonucleotides of several eukaryotic initiator tRNAs (20). In all these cases, this material also appears almost immediately upon addition of SVP, before the exonucleolytic digestion has progressed to shorter fragments (20).


A
B
B
B
C
D
pC
U
C
A
U
^{6}A
A
C
C
C
G
2
1

E. Analysis of Longer Nucleotide Sequences by Partial Digestion of [^{32}P]-End Group Labeled Material with Nuclease P_1

To complete an RNA sequence, it is necessary to order the shorter fragments obtained by T_1 RNase or pancreatic RNase digestion into a unique total sequence. We have found the use of partial digestion with nuclease P_1 of high molecular weight end-group labeled RNA to be a very useful approach to this end. As will be illustrated below for TYMV RNA, the 3'- and 5'-termini of molecules as large as viral genomes can be sequenced directly by two-dimensional homochromatography if they can be cleanly end-group labeled; sequences of over 20 nucleotides can often be determined from a single two-dimensional analysis. In contrast to most analogous methods previously described for DNA sequencing (148,149), prior purification of terminal polynucleotide fragments is not necessary, making this system both faster and more efficient of sample.

For tRNA, application of this methodology involves preparation of 5'-[^{32}P]-labeled tRNA with polynucleotide kinase and 3'-[^{32}P]-labeled tRNA with tRNA nucleotidyl transferase (see Methods). The 5'- and 3'-labeled tRNAs are purified by gel electrophoresis under denaturing conditions (7 M urea). They are then partially degraded under mild conditions with nuclease P_1 (Fig. 12A) and analyzed directly by two-dimensional homochromatography. Terminal sequences of at least 15 nucleotides can usually be determined in a single experiment.

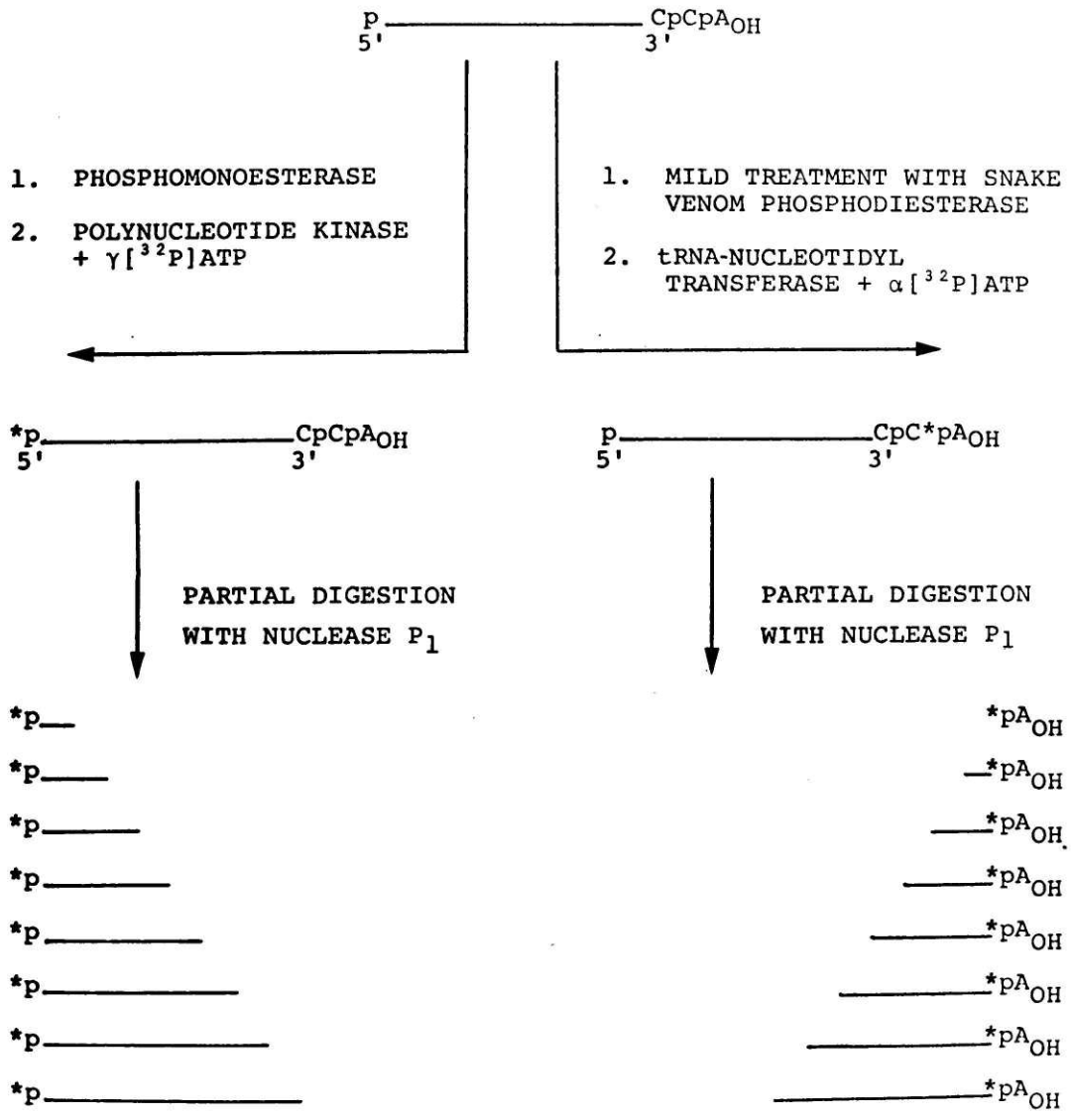


Figure 12A. Scheme for the use of nuclease P_1 for sequence analysis of end-labeled tRNA. Mobility shift analysis is essentially identical for two-dimensional homochromatographic patterns of digests of 5'-labeled or 3'-labeled RNA.

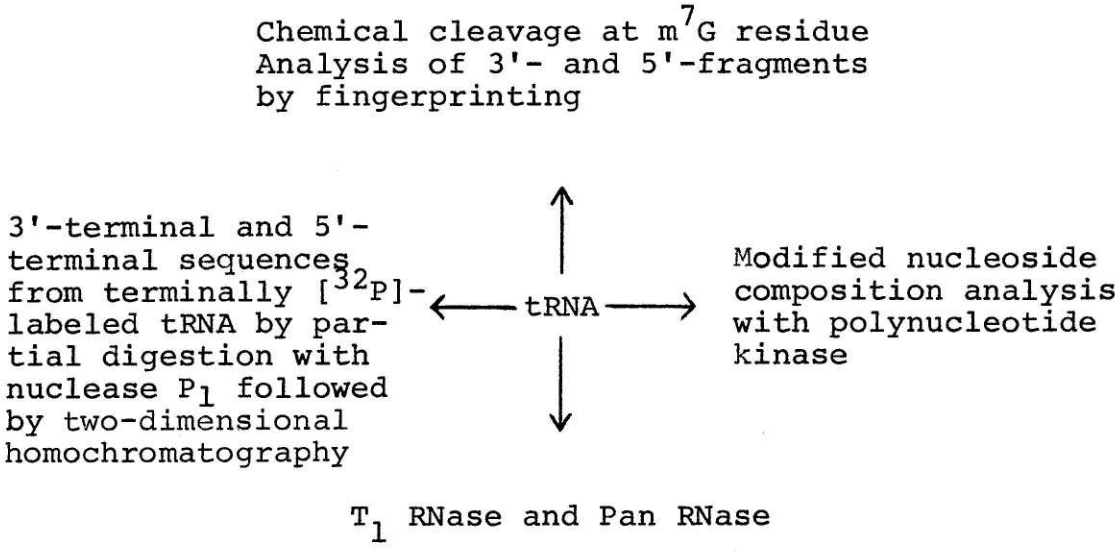


Figure 12B. Scheme for the complete sequence analysis of tRNA by the methods described.

Large fragments from internal positions in an RNA molecule may be obtained by partial degradation of the molecule with T_1 RNase or pancreatic RNase[1]. These fragments, too, can be 5'-terminally labeled with [^{32}P], and can then be fractionated by two-dimensional gel electrophoresis (116; Methods). They can then be analyzed both by fingerprinting, to determine their content of previously sequenced shorter fragments, and by direct 5'-terminal sequence analysis with nuclease P_1. As is discussed below (Results IIIA), this approach has provided sufficient data in the case of Euglena gracilis chloroplastic $tRNA^{Phe}$ to allow determination of the complete nucleotide sequence. It is anticipated that this approach will be generally sufficient to determine RNA sequences of molecules of tRNA size. An overall scheme for the sequencing of tRNA molecules, in particular, by [^{32}P]-end group labeling techniques is summarized in Fig. 12B.

Despite its many important applications, several limitations on the use of nuclease P_1 and two-dimensional homochromatography in sequence analysis should be noted. First, nuclease P_1 is not entirely random in its selection of cleavage sites (105). In addition to slow cleavage on the 3'-side of certain modified nucleosides, including t^6A, Ψ, and O-methylated nucleosides (data not shown), the enzyme

---

[1]Another method of tRNA cleavage into large fragments, involving chemical cleavage adjacent to the m^7G residue (16), will not be discussed here.

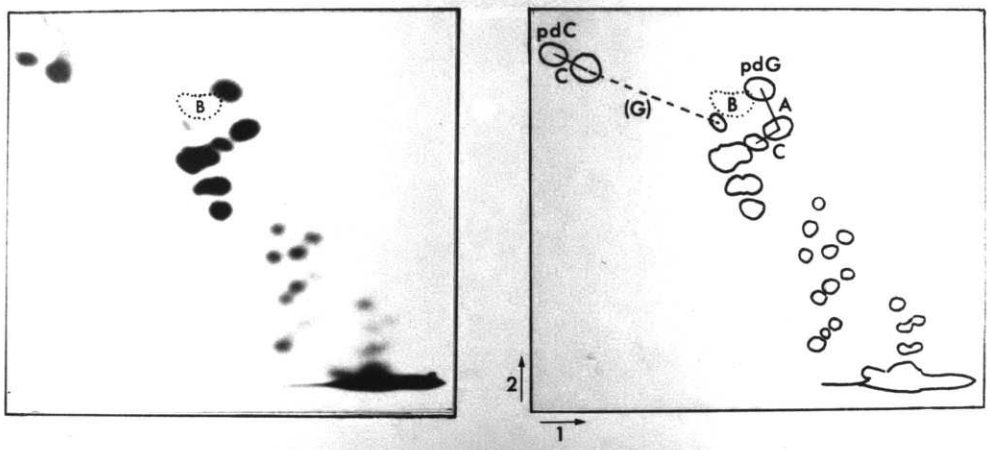
is somewhat slowed by stretches of pyrimidines, especially C (cf. Fig. 11B). A two-dimensional homochromatographic TLC pattern of a partial digest produced exclusively with nuclease P_1 may therefore contain sequence "gaps" (cf. Fig. 11B); these must be filled either by knowledge of the sequences of component shorter fragments previously analyzed by methods discussed above, or, in the case of shorter oligonucleotides, by analysis of SVP-generated partial digests (cf. Fig. 11). With regard to the limitations of two-dimensional homochromatography, it should be noted that the mobility shift analysis degenerates in the case of long, uninterrupted (G,U)-stretches; after 6 or 7 residues, such fast-moving compounds no longer obey the characteristic first-dimensional mobility shifts (data not shown), and sequence analysis is not possible. Possibly because of their reliance on analysis of relatively (A,C)-rich reference compounds, Wu and collaborators (146) seem to have ignored this fact.

Finally, it was of interest to examine whether nuclease P_1 could also be used in DNA sequencing. Accordingly, a restriction fragment (sequence unknown) from the DNA genome of phage ø80pSUIII was 5'-terminally labeled with [^{32}P] using polynucleotide kinase (as described in Methods) and repurified by polyacrylamide gel electrophoresis in 7 M urea (as described above for RNA). This DNA fragment, designated DlAZ1 by A. Landy (in whose laboratory it was prepared; Materials), contained a cut on one side by the restriction

enzyme Hemophilus influenzae d II (Hind II) and a cut on the other side by the restriction enzyme Hemophilus aegyptius III (Hae III); its expected 5'-end sequences are therefore purine-A-C (150) and C-C (151), respectively. Upon treatment with nuclease P_1 and analysis by two-dimensional homochromatography, using procedures identical to those used previously with RNA, the 5'-terminally labeled DNA yielded the pattern shown in Fig. 13. As can be seen from the figure, analysis of the end sequences is consistent with the recognition sites of the restriction nucleases used in the preparation of the fragment. 5'-end group analysis (Methods) showed an almost equal distribution of radioactivity in pdC and pdG (data not shown), consistent with the equal presence of the two DNA strands.

The DNA analyzed in Fig. 13 was in a denatured condition (having been recovered from a denaturing polyacrylamide gel) during partial digestion with nuclease P_1. It has since been found that native double-stranded DNA, too, is suitable for analysis by partial digestion with nuclease P_1; however, the reaction works best if performed at 37°C rather than at room temperature (data not shown). However, many restriction enzymes leave free single-stranded 5'-end regions after cleavage ("sticky ends"); these single-stranded regions are digested more rapidly by nuclease P_1 than double-stranded regions. DNA fragments ending in single-stranded 5'-ends are best denatured (e.g., by boiling) before sequence analysis with nuclease P_1.

Figure 13. Analysis by two-dimensional homochromatography of a partial digestion, with nuclease P_1, of the 5'-[^{32}P]-labeled restriction nuclease-generated DNA fragment DlAZl (for details see text). Pattern shown is an autoradiogram; B (surrounded by dotted circle) is xylene cyanole dye marker. Note the two overlapping patterns that arise from the two 5'-[^{32}P]-labeled ends of this double-stranded DNA.


B
pdC
C
(G)
pdG
B
A
C
2
1

To use nuclease P_1 for DNA sequence analysis, it would be necessary to prepare DNA labeled at only one 5'-end. This could be accomplished by digesting a fragment already labeled at both 5'-ends with another restriction enzyme that cuts the fragment internally; the two "halves" of the fragment could then be purified by gel electrophoresis and sequenced. Alternatively, the two 5'-terminally labeled strands of a DNA duplex could be separated from each other and analyzed individually. The use of nuclease P_1 may thus provide a useful method to supplement other DNA sequencing techniques.

II. Studies on the Sequence of a 3'-Terminal Fragment of TYMV RNA

For the purposes of this discussion, the 3'-terminal 4.5S RNA fragment prepared from TYMV RNA, as described under Methods, above, will be referred to as the TYMV RNA 3'-fragment. This work was conducted in collaboration with A. Prochiantz and A.L. Haenni of the University of Paris, France.

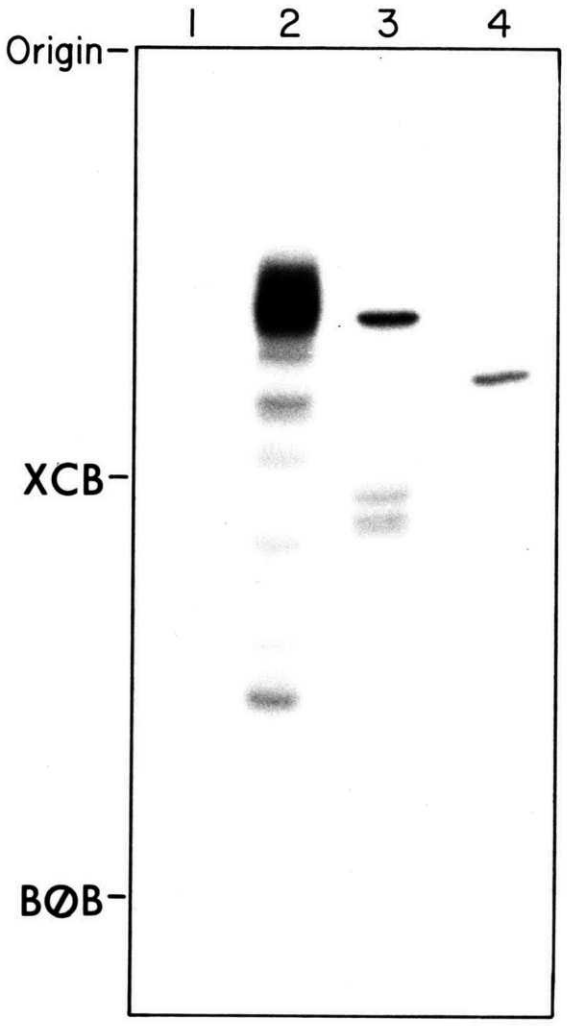
A. Estimation of the Size and Purity of the TYMV RNA 3'-Fragment

The TYMV RNA 3'-fragment (0.5 μg) was labeled at its 5'-end with [^{32}P] using polynucleotide kinase (as described in Methods), and the [^{32}P]-labeled RNA was then electrophoresed in a 12% or a 15% polyacrylamide slab gel in 7 M urea

alongside [^{32}P]-labeled marker RNAs of known size. The markers used were *in vivo* [^{32}P]-labeled *E. coli* $tRNA_I^{Tyr}$ precursor (129 nucleotides long), and 3'-[^{32}P]-labeled *E. coli* 5S RNA (120 nucleotides long), and 3'-[^{32}P]-labeled brewer's yeast $tRNA^{Phe}$ (76 nucleotides long). Based on its electrophoretic mobility relative to the marker RNAs, the TYMV RNA 3'-fragment was estimated to have a size of ca. 112 ± 3 nucleotides in a 15% gel (data not shown), and in a 12% gel (Fig. 14; track 3). The autoradiogram of Fig. 14 also indicates the relatively high degree of purity of this fragment preparation (approximately 80%), and it was the preparation used for most of the experiments discussed below (possible sequence contaminants are discussed below).

Although the TYMV RNA 3'-fragment shown in Fig. 14 was treated with phosphatase before 5'-labeling, it was found that untreated fragment accepted a 5'-[^{32}P] equally well (extent of phosphorylation 75-100%); tRNA not previously treated with phosphatase, however, accepted no [^{32}P]. This indicates that the fragment as purified here contains a free 5'-hydroxyl end. When the predominant radioactive band present in the 5'-[^{32}P]-labeled TYMV RNA 3'-fragment (Fig. 14, track 3) was extracted and then digested with T_2 RNase to mononucleotides the 5'-[^{32}P] was found by one-dimensional TLC (solvent (a)) to be 83% pAp, 15% pCp and 2% pUp, thus indicating a slight heterogeneity at the 5'-end of this fragment. Both the presence of a 5'-hydroxyl end and the slight heterogeneity at the 5'-end were observed in several different preparations of the TYMV

Figure 14. Autoradiogram of polyacrylamide gel electrophoresis of the 5'-[^{32}P]-labeled TYMV RNA 3'-fragment along with RNA markers of known length. Electrophoresis was in a 12% polyacrylamide gel slab containing 7 M urea. 1) E. coli $tRNA_I^{Tyr}$ precursor (129 nucleotides long); 2) E. coli 5S RNA (120 nucleotides long); 3) TYMV RNA fragment; 4) brewer's yeast $tRNA^{Phe}$ (76 nucleotides long). XCB, xylene cyanole; BØB, bromephenol blue.

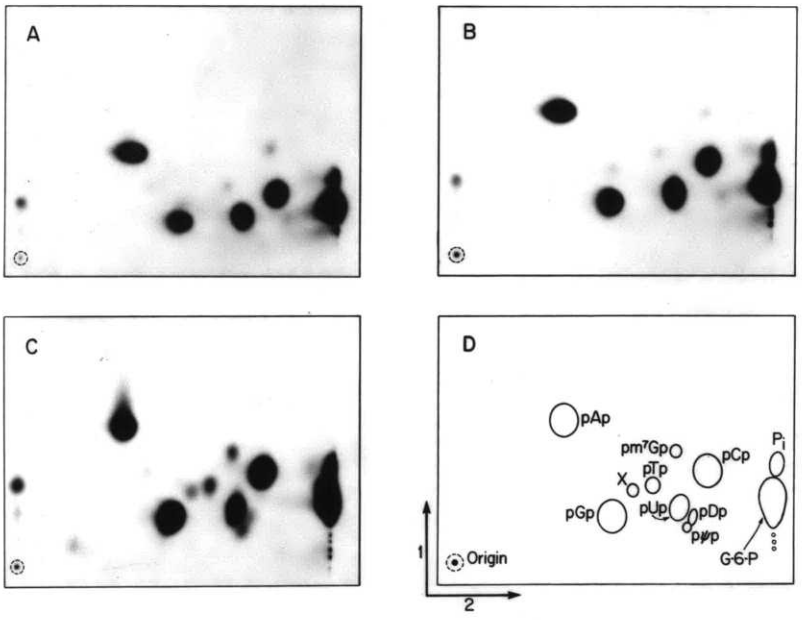

1
2
3
4
Origin
XCB
BØB

RNA 3'-fragment. Since cleavage by "RNase P" yields an RNA containing a 5'-phosphate group rather than a 5'-hydroxyl (79), and, particularly, in view of the slight heterogeneity at the 5'-end of the TYMV RNA 3'-fragment, it is possible that the enzyme involved in the production of this fragment is not "RNase P" but another nuclease or nucleases which have co-purified with "RNase P". This matter is treated more fully below (Discussion ID).

B. Analysis for the Presence of Modified Nucleosides in the TYMV RNA 3'-Fragment

The possible presence of modified nucleosides was also investigated by in vitro [^{32}P]-labeling. The TYMV RNA 3'-fragment was first digested with T_2 RNase and then treated with polynucleotide kinase in the presence of γ[^{32}P]ATP. Analysis was by two-dimensional TLC in solvents (a) and (c), as described above (Fig. 15). It can be seen that the pattern obtained from the TYMV RNA 3'-fragment (Fig. 15A) is essentially identical to that obtained from the tetranucleotide A-U-C-Gp (Fig. 15B) subjected to the same treatment. An analysis of E. coli $tRNA_f^{Met}$ (Fig. 15c and 15d) is included here to emphasize, by contrast, that modified nucleosides, when present, are easily detected in this system. Based on Fig. 15 and on many further observations on the behavior of modified nucleosides in this type of analysis (cf. Fig. 4, above), we may conclude that the TYMV RNA fragment contains no m^7G, T, D, Ψ, m^1G or m^1A. Although all modifications cannot be specifically excluded, we feel that the fragment is most likely free of modified nucleosides.

Figure 15. Analysis of the TYMV RNA 3'-fragment for the presence of modified nucleosides. Autoradiograms of two-dimensional thin layer chromatography patterns of [^{32}P]-labeled nucleoside-3',5'-diphosphates obtained from (for details see Methods) A) TYMV RNA 3'-fragment; B) A-U-C-Gp; and C) E. coli $tRNA_f^{Met}$. D) is a replica of C) indicating identification of spots. The spot labeled X was later identified as pCm-Up (Fig. 2). G-6-p is glucose-6-phosphate. For discussion of artifactual radioactive spots at left of each plate see Legend, Fig. 2.


A
B
C
D
pAp
pm⁷Gp
pCp
Pi
pTp
X
pGp
pUp
pDp
pψp
G-6-P
Origin
1
2

C. Analysis of Oligonucleotides Present in Complete T_1 RNase and Pancreatic RNase Digests of the TYMV RNA 3'-Fragment

Fig. 16 shows the fingerprints of complete T_1 RNase (Fig. 16A) and pancreatic RNase (Fig. 16B) digests of the TYMV RNA 3'-fragment. Fig. 17 shows the fingerprint of a complete T_1 RNase digest using PEI-cellulose chromatography instead of DEAE-paper ionophoresis in the second dimension (see Methods). The total nucleotide sequences and molar yields of all the radioactive oligonucleotides thus obtained are listed in Table V. The sequences indicated in the table were determined by partial digestion with SVP and analysis by one-dimensional DEAE-paper ionophoresis and/or by two-dimensional homochromatography (see above; procedural details in Methods). A detailed description of the analysis of several of these sequences follows.

Spot T2 was missing from the standard T_1 RNase fingerprint of Fig. 16A, but was present in 0.5 molar yield in a PEI-cellulose TLC fingerprint (Fig. 17) prepared from the same [^{32}P]-labeled digest; in other fingerprints which used DEAE-paper ionophoresis it was present in 0.1-0.5 molar yield. Spot T9 is also difficult to discern in Fig. 16A, where it has streaked badly in the second dimension, yet can be seen quite clearly in fingerprints which have used PEI-cellulose TLC (Fig. 17). Both spots T2 and T9 were therefore prepared exclusively from PEI-cellulose fingerprints.

Figure 16. Autoradiograms of 5'-[^{32}P]-labeled oligonucleotides obtained from T_1 RNase (A) and pancreatic RNase (B) digests of the TYMV RNA 3'-fragment as analyzed by two dimensional ionophoresis. In (A), spot 2 is faint and was not numbered, and spot 9 streaked in the second dimension; both these spots appear more clearly in the fingerprint of Fig. 17, below, which was prepared from an aliquot of the same polynucleotide kinase incubation. See also legend to Table V. The spot numbered 17 in (B) is artifactual and was correlated with the particular preparation of γ[^{32}P]ATP used in this case. Several other faint spots in B), not numbered, were not present in other pancreatic RNase digests of the TYMV RNA 3'-fragment (data not shown), and were not studied. ATP is an occasional contaminating product due to its incomplete conversion to glucose-6-phosphate. B surrounded by dots, xylene cyanole blue dye marker.

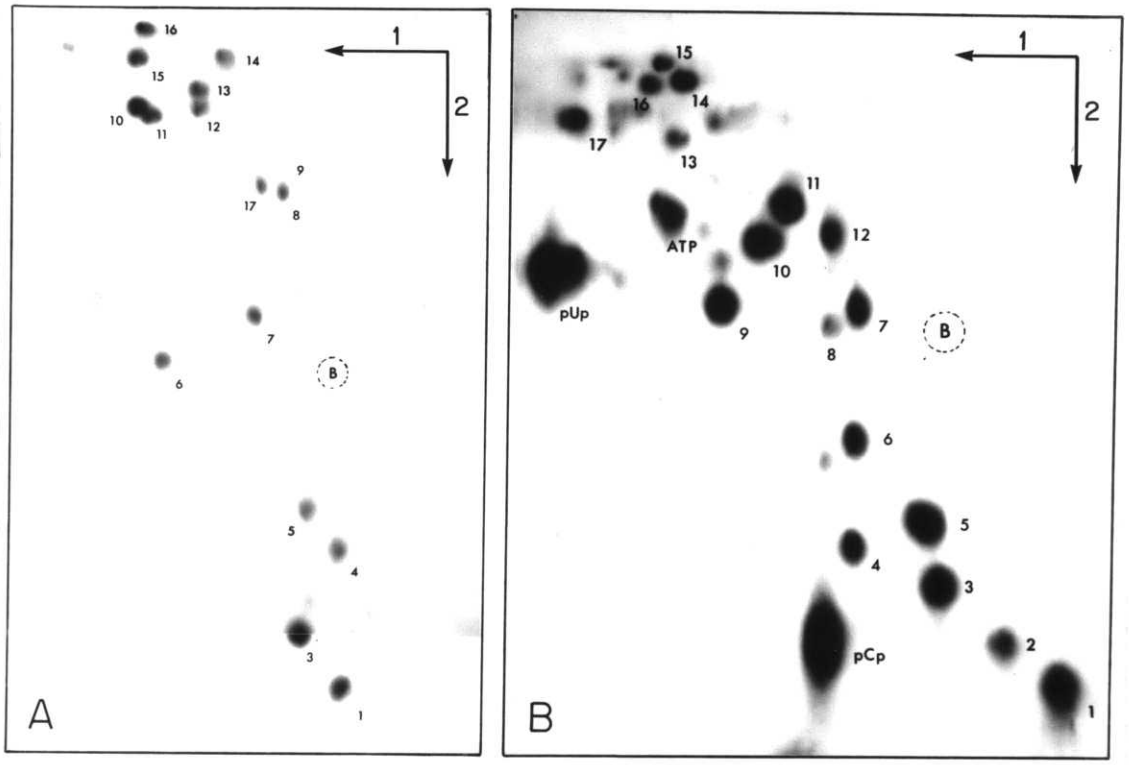

A
16
15
14
13
10
11
12
9
17
8
7
6
B
5
4
3
1
1
2
B
15
16
14
17
13
11
ATP
12
10
pUp
9
7
8
B
6
5
4
3
pCp
2
1
1
2

Figure 17. Autoradiogram of 5'-[^{32}P]-labeled oligonucleotides obtained from a T_1 RNase digest of the TYMV RNA 3'-fragment. The [^{32}P]-labeled digest used was the same as that of Fig. 16A. First dimension was ionophoresis on cellulose acetate at pH 3.5 as in Fig. 16; second dimension was thin layer chromatography on a PEI-cellulose plate as described in the text. The spot labeled X was a complex mixture of [^{32}P]-containing material that was assumed to be artifactual and was not analyzed further. B surrounded by dots is xylene cyanole blue dye marker; glucose-6-p is glucose-6-phosphate and P_i is inorganic phosphate.

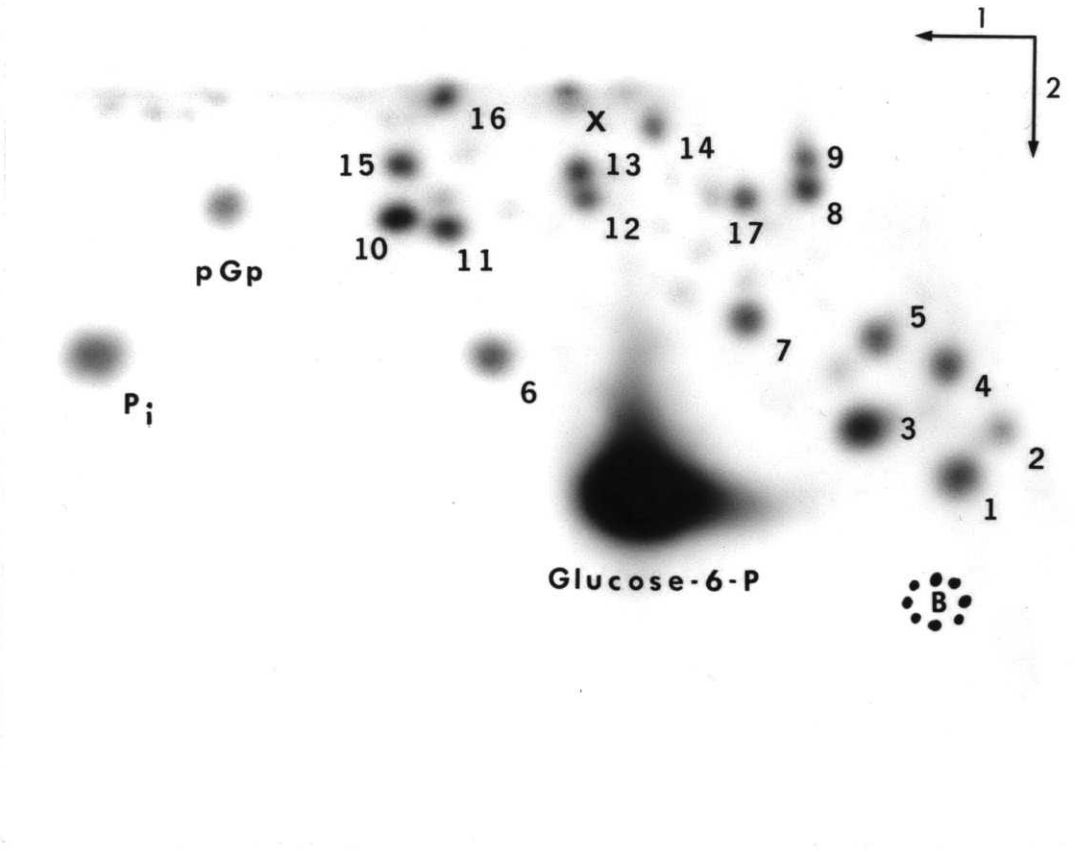

1
2
16
X
14
15
13
9
8
12
17
10
11
pGp
7
5
6
4
P_i
3
2
1
Glucose-6-P
B

Table V. Oligonucleotide Sequences and Molar Yields Present in Fingerprints (Fig. 16A and B) of T_1 RNase (A) and Pancreatic RNase (B) Digests of the TYMV RNA 3'-Fragment.

---

*Determined by Cerenkov counting of spots excised from fingerprints of two independent preparations of the TYMV RNA 3'-fragment.

+It should also be noted that intact TYMV RNA terminates with C at its 3'-end (67). The presence of a 3'-terminal A in this sequence (spot T2) is due to the fact that the TYMV RNA was aminoacylated prior to purification of the fragment. The variable yield of the 3'-end sequence is believed to be due to partial removal of the phenoxyacetylvalyl moiety in the course of RNase digestion, 5'-labeling, and fingerprinting.

‡This sequence was present in near molar amounts only in the cleanest fragment preparation, and is assumed to be the 5'-end sequence generated from the most susceptible "RNase P" cleavage point.

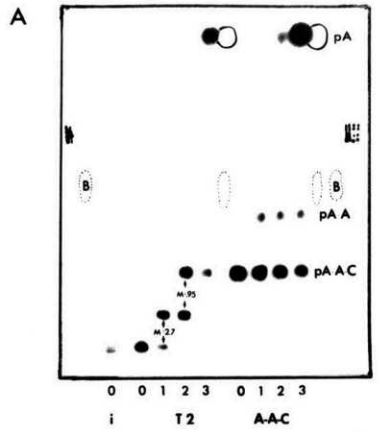
	Spot No.	Sequence	Molar Ratio*
A.	T1	C-G	1.4
	T2	A-A-C-C-A	0.1-0.5[+]
	T3	A-G	2.7
	T4	C-C-A-G	1.0
	T5	A-C-A-G	0.98
	T6	U-G	1.4
	T6	C-U-C-G	1.2
	T8	C-A-A-C-U-C-C-C-G	0.70
	T9	U-C-C-C-C-A-C-A-C-G	1.0
	T10	U-U-A-G	2.2
	T11	U-C-U-G	0.97
	T12	U-C-A-U-C-G	0.82
	T13	A-U-A-A-U-C-G	0.92
	T14	C-C-C-C-U-C-U-U-C-C-G	0.91
	T15	U-U-C-U-C-G	1.2
	T16	A-U-C-U-U-U-A-A-A-A-U-C-G	0.81
	T17	A-C-C-U-A-A-G	0.1-0.67[++]
B.	A1	A-C	2.2
	A2	A-A-C	0.8
	A3	G-C	2.4
	A4	A-U	0.91
	A5	2A-G-C + 1G-A-C	2.5
	A6	A-A-U	1.1
	A7	A-A-A-A-U	0.74
	A8	G-G-C	0.20
	A9	G-U	1.5
	A10	1A-G-U + 1G-A-U	2.4
	A11	1A-A-G-U + 1A-G-A-U	2.2
	A12	G-G-A-A-C	0.88
	A13	G-G-U	0.28
	A14	G-A-G-G-U	0.69
	A15	G-A-G-G-G-U	0.33
	A16	G-G-G-U	0.40

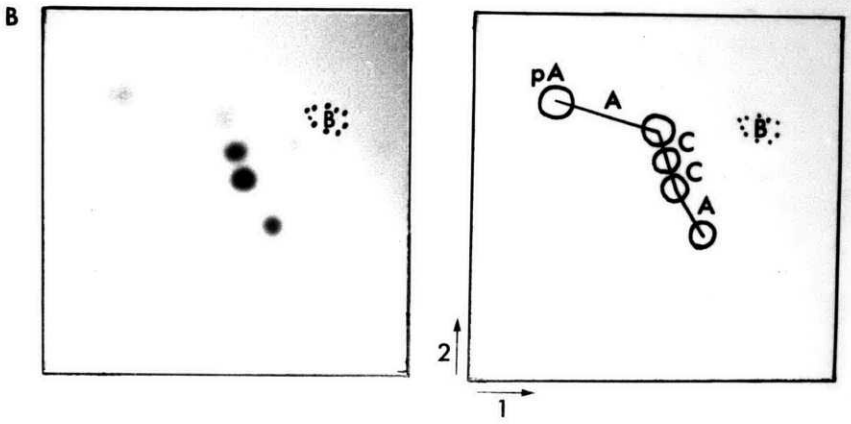
The sequence of oligonucleotide T2 was determined both by DEAE-paper ionophoresis (Fig. 18A) and by two-dimensional homochromatography (Fig. 18B). These two analyses established conclusively that the sequence of this oligonucleotide is A-A-C-C-A, showing it to be the 3'-end of the fragment (this sequence does not agree with the tentative 3'-end sequence assigned by Litvak et al. (152)). We believe that the variable yield of spot T2 (Table Va and above) is due to the incomplete and variable removal of the phenoxyacetylvalyl moiety in the course of RNase digestion, 5'-labeling and fingerprinting.

The sequences of di-, tri- and tetranucleotides were determined by DEAE-paper ionophoresis (data not shown); A. Prochiantz participated in the analysis of shorter oligonucleotide sequences from the pancreatic RNase fingerprint (Fig. 16B).

Of the shorter oligonucleotides, T11 (Table Va) was of particular interest. Its mobility in the fingerprint (Fig. 16A) is consistent with the nucleotide composition (U_2, C, G), it has a U 5'-end group (data not shown), and partial digestion with SVP, when analyzed by one-dimensional homochromatography (cf. Fig. 7), shows it to be a tetranucleotide (data not shown). It is therefore a potential U-U-C-G sequence, as might be generated by T_1 RNase cleavage of loop IV of a typical tRNA (Fig. 1). However, analysis by one-dimensional DEAE-paper ionophoresis at pH 3.5 and at pH 1.9

Figure 18. Autoradiograms of analyses of the sequence of oligonucleotide T2 (Fig. 16A, Table Va) by A) one-dimensional ionophoresis at pH 3.5 alongside the oligonucleotide of known sequence *pC-A-A-C-C-A (i), and a partial digest of the known oligonucleotide *pA-A-C; B) two-dimensional homochromatography. B circled by dots in A) and B) is xylene cyanole dye. The method of obtaining partial digests of oligonucleotides with SVP and of selecting time point aliquots of the digestion reaction for analysis is described in Methods and Results I. (Note, paper electrophoresis and thin layer chromatography patterns have not been reduced to the same scale.)


A
pA
B
B
pA A
pA A C
M 95
M 27
0 0 1 2 3 0 1 2 3
i
T 2
A-A-C


B
pA
A
C
C
A
2
1

(data not shown), and analysis by two-dimensional homochromatography (Fig. 19), unambiguously indicated the sequence to be U-C-U-G.

Two of the pancreatic RNase fingerprint oligonucleotides, A5 and A10 (Table Vb), contained both A and G 5'-end groups; this indicated a mixture of sequences. The ratio of the sequences present was determined, in each case, by the ratio of [^{32}P] cpm in the A and G end group spots recovered from TLC plates. The sequences in A5, A-G-C and G-A-C, and the sequences in A10, A-G-U and G-A-U, could be inferred from the mobilities of these spots in the fingerprint (Fig. 16B). They were confirmed by partial SVP digestion and analysis by one-dimensional paper ionophoresis at pH 3.5; in each case, two dinucleotide spots, A-G and G-A, were slightly resolved (data not shown).

The 5'-end group of oligonucleotide A11 is A. However, when a partial SVP digest of A11 is analyzed by ionophoresis on DEAE-paper at pH 3.5, two dinucleotide spots appear, A-A and A-G (Fig. 20A). These are followed by two trinucleotide spots, the faster of which comigrates with an A-A-G marker compound (kindly supplied, from a T_1 RNase fingerprint of Neurospora crassa $tRNA_i^{Met}$, by Dr. A. Gillum) (Fig. 20A). It was confirmed by two-dimensional homochromatography (Fig. 20B) that this oligonucleotide contained a mixture of two compounds, A-A-G-U and A-G-A-U. It was also inferred from the equal intensities of the two di- and the two

Figure 19. Autoradiogram of analysis by two-dimensional homochromatography of a partial SVP digest of oligonucleotide T11 (Fig. 16A; Table Va). The sequence deduced is *pU-C-U-G.

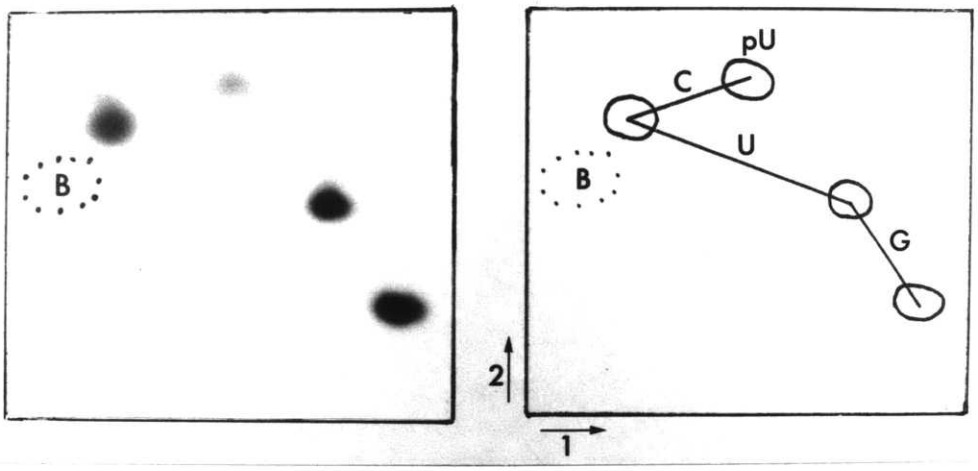
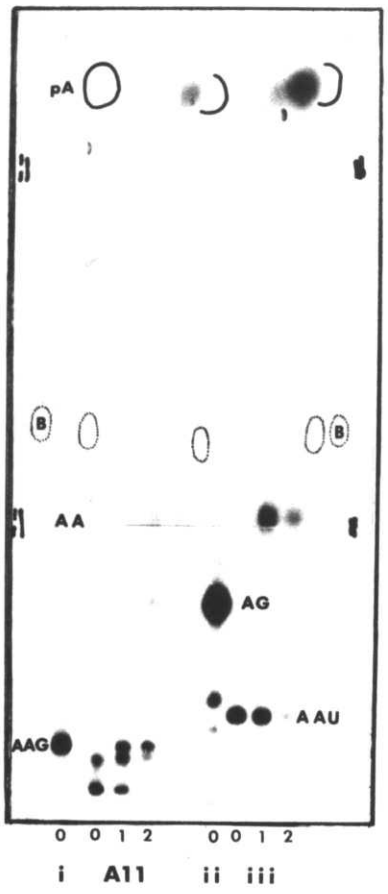
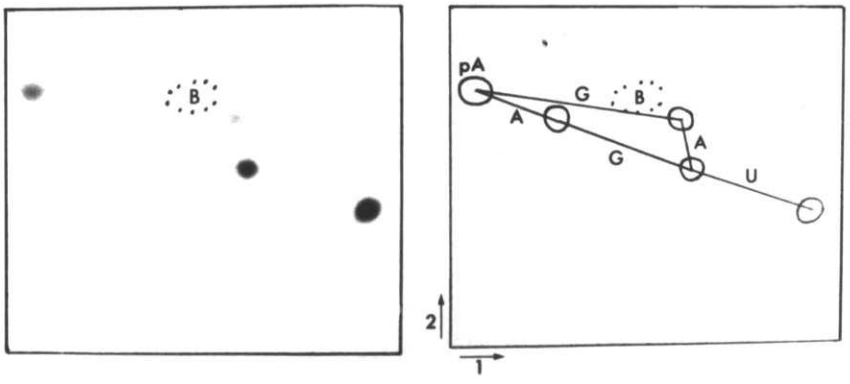

B
pU
C
U
B
G
2
1

Figure 20. Autoradiograms of analyses of the sequence(s) of oligonucleotide(s) A11 (Fig. 16B; Table Vb) by A) one-dimensional ionophoresis at pH 3.5 alongside the oligonucleotides of known sequence i) *pA-A-G, ii) *pA-G and a partial SVP digestion of the known oligonucleotide iii) *pA-A-U; B) two-dimensional homochromatography. B circled by dots is xylene cyanole dye. The sequence deduced for A11 is (Table Vb) an equimolar mixture of *pA-A-G-U and *pA-G-A-U. (Note, paper electrophoresis and thin layer chromatography patterns have not been reduced to the same scale.)

A



B



trinucleotide spots in Fig. 20A that these two sequences were present in equimolar amounts.

Sequences longer than tetranucleotides were generally determined by combining the results of analyses of partial SVP digestions by one-dimensional DEAE-paper ionophoresis and by two-dimensional homochromatography. Analysis of pancreatic RNase fingerprint oligonucleotides was particularly straightforward, as they contain only one 3'-terminal pyrimidine residue. Where only G and A need be unambiguously differentiated, homochromatography provides the simplest means of anlaysis. This is illustrated for the sequence analysis of Al4 and Al5 (Fig. 21). It can be seen from Fig. 21 that these two similar sequences, G-A-G-G-U and G-A-G-G-G-U, respectively (Table Vb), can be "read" directly and unambiguously by mobility shift analysis. In RNA, mobility shifts for A and G in the cellulose acetate (pH 3.5) dimension can generally be clearly distinguished; this is not necessarily true for DNA, however (145,146). The sequences of Al4 and Al5 were corroborated by one-dimensional DEAE-paper ionophoresis at pH 1.9 (data not shown).

The ionophoretic analysis of oligonucleotides Tl5, Tl2, Tl3 and Tl6 shown above (Figs. 8, 9) were all confirmed by two-dimensional homochromatography. Fig. 22 shows the analyses of Tl5 and Tl6; it should be noted that, unlike the one-dimensional ionophoretic analyses of these sequences discussed above (section I), there is no ambiguity whatsoever

Figure 21. Autoradiograms of analyses by two-dimensional homochromatography of the sequences of A) oligonucleotide A14 (Fig. 16B, Table Vb) and B) oligonucleotide A15 (Fig. 16B, Table Vb). B circled by dots is xylene cyanole dye.

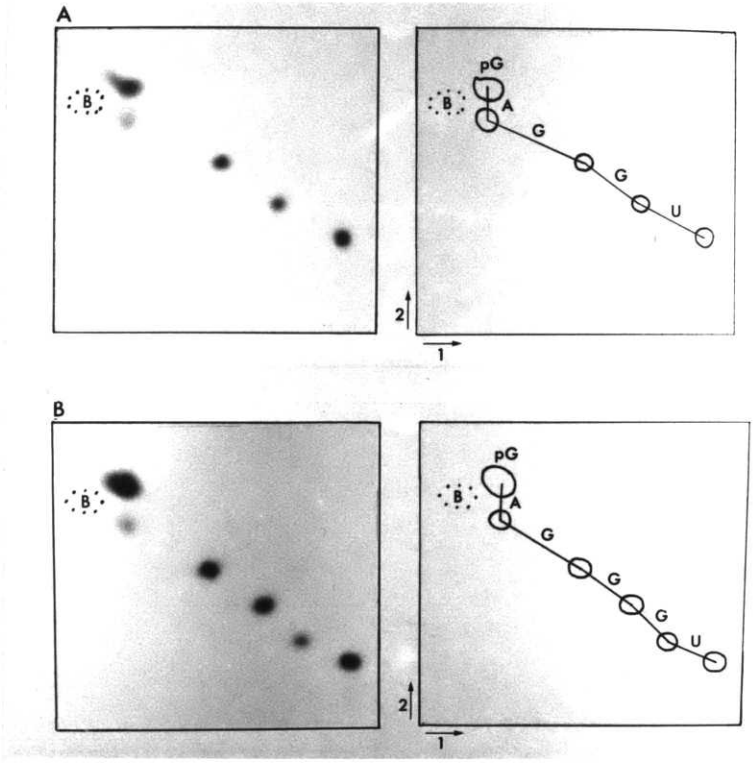
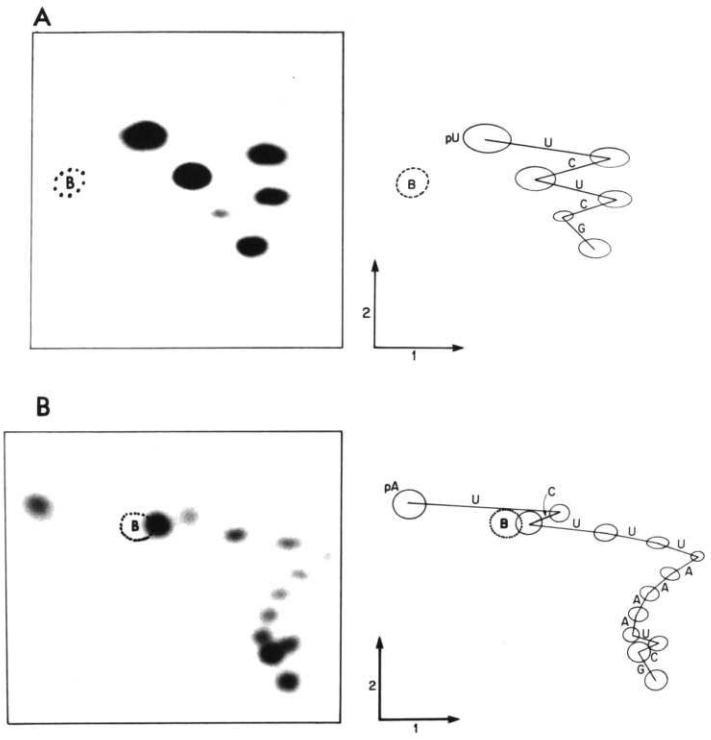

A
B
pG
A
G
G
U
2
1
B
B
pG
A
G
G
G
U
2
1

Figure 22. Autoradiogram of analysis by two-dimensional homochromatography of the sequence of A) oligonucleotide T15 (Fig. 16A, Table Va) and B) oligonucleotide T16 (Fig. 16A, Table Va). B circled by dots is xylene cyanole dye. Compare Figs. 7 and 8.

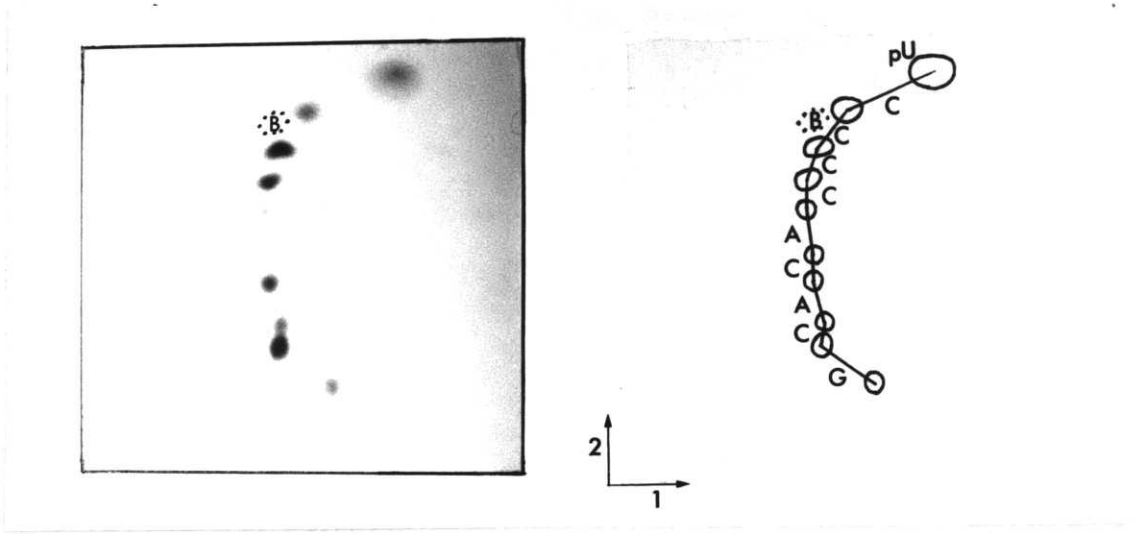

A
B
pU
U
C
U
C
G
2
1
B
pA
U
C
U
U
U
A
A
A
A
U
U
C
G
2
1
B

in the two-dimensional homochromatography analyses. The analysis of T16 in Fig. 22 also illustrates the suitability of two-dimensional homochromatography for analysis of longer sequences; the ononphoretic data are thus both corroborated and extended.

Fig. 23A shows the analysis, by two-dimensional homochromatography, of the sequence of oligonucleotide T9; it is included here to illustrate the clear distinction of purine and pyrimidine mobility shifts in this (A,C)-rich sequence by the homochromatography dimension. These distinctive mobility shifts (Fig. 23A) allow an unambiguous sequence determination. Fig. 23B shows the analysis, by two-dimensional homochromatography, of the sequence of oligonucleotide T14, and illustrates the behavior of sequences of C's at the 5'-terminal end of an oligonucleotide, in the cellulose acetate (pH 3.5) dimension. As can be seen from the figure, pC-C runs faster than pC, and pC-C-C faster than pC-C, etc., as opposed to the more usual direction of C mobility shifts (Fig. 10A). This phenomenon has been discussed above (Section I). The 5'-proximal sequence of T14 was confirmed by one-dimensional ionophoresis on DEAE-paper at pH 3.5 alongside a partial digest of the marker compound pC-C-C-C-A-C-G (Fig. 24). Once thus established unambiguously, the sequence of T14 could be used as a two-dimensional homochromatography reference pattern for comparison with the patterns of unknown sequences (cf., the DNA sequence pdC-C...

Figure 23. Autoradiograms of analyses by two-dimensional homochromatography of the sequences of A) oligonucleotide T9 (Fig. 16A; Table Va) and B) oligonucleotide T14 (Fig. 16A, Table Va). B circled by dots, xylene cyanole dye.

A



B

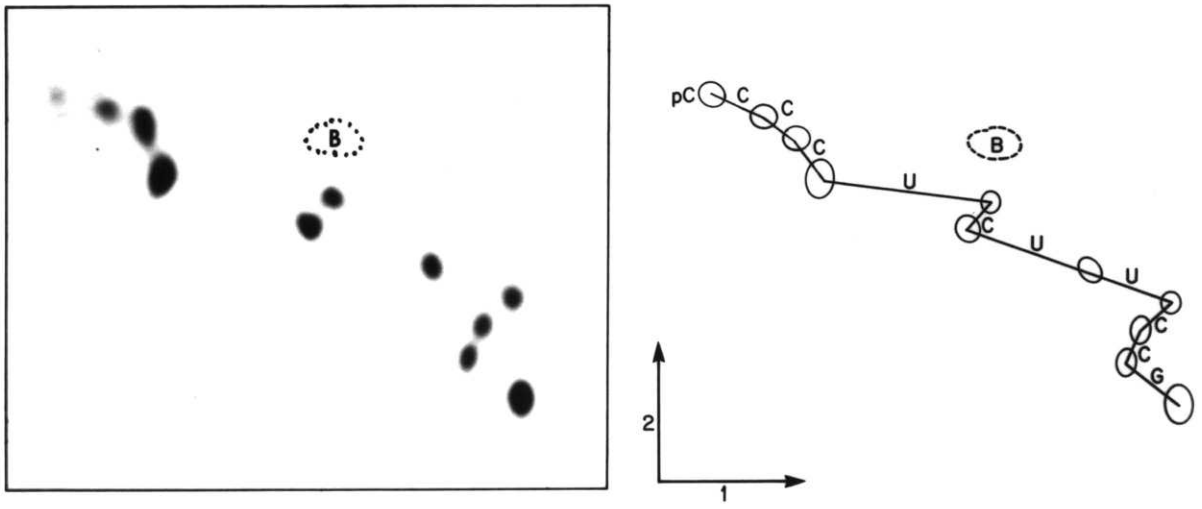
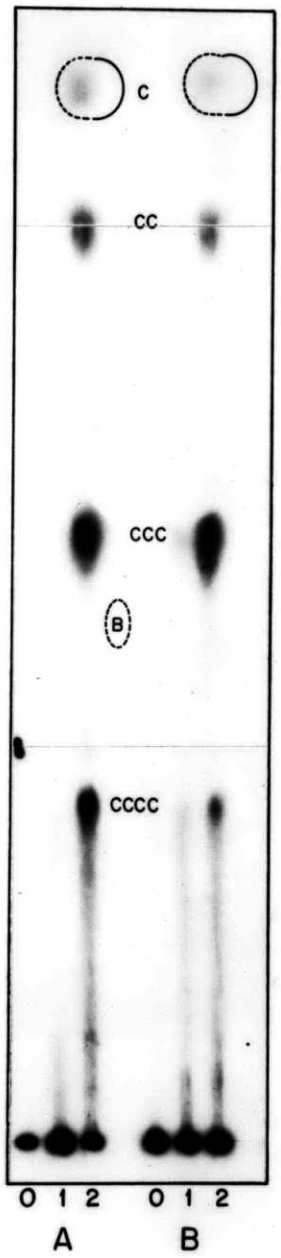


Figure 24. Autoradiogram of a partial SVP digest of spot T14 of Fig. 16A as analyzed by ionophoresis on DEAE paper at pH 3.5. Three sets of pooled time points from partial digestion of A) oligonucleotide T14 of Fig. 16A and B) an oligonucleotide of known sequence, *pC-C-C-C-A-C-G, used here a a marker for the 5'-proximal sequence *pC-C-C-C. B surrounded by dots is xylene cyanole dye.


C
CC
CCC
B
CCCC
0 1 2 0 1 2
A B

of Fig. 13, the "reading" of which in fact relied on a knowledge of T14 (Fig. 23B)).

Unfortunately, quantitation of oligonucleotide molar yields was not as straightforward as sequence determination. Table V indicates that besides T2 (discussed above), oligonucleotide T17 is also recovered in low and variable yield. T14 was present in near-molar amounts only in the cleanest fragment preparation (Fig. 14), and is assumed to be the 5'-end sequence generated from the most susceptible "RNase P" cleavage point. Table Vb indicates that the molar yields recovered of spots A8, A13 and A15 are all below 40%. Sequence A15 was later shown by an independent experiment (Fig. 25 and below) to be near the 3'-end of the fragment. Whether A8 and A13 are genuine components of the TYMV RNA 3'-fragment which are recovered in low yield, or whether they arise from contaminants in the TYMV RNA 3'-fragment is not known (note that the purity of this fragment preparation is only approximately 80% (Fig. 14)).

D. Sequence of the 5'- and 3'-Terminal Regions of the TYMV RNA 3'-Fragment

As discussed above (Section I), partial digestion of 5'- or 3'-labeled tRNAs with nuclease P_1 followed by two-dimensional homochromatography can provide substantial sequence information for the 5'- and the 3'-terminal regions of these tRNAs. This procedure has proved very usefuly in analysis of the TYMV RNA 3'-fragment.

TYMV RNA, which ends in ...C-C_{OH} as extracted from the virion (67,152), was treated with purified E. coli tRNA nucleotidyltransferase and $\alpha[^{32}P]$ATP to obtain 3'-$[^{32}P]$-labeled material ending in ...C-C-$[^{32}P]$-A_{OH}. This material was then digested with nuclease P_1 under mild conditions to generate a random population of $[^{32}P]$-labeled partial degradation products which derive from the 3'-end (see Methods). Fig. 25 shows an autoradiogram of a two-dimensional homochromatography pattern of such a digest. Based on the known sequence of the 3'-terminal oligonucleotide as A-A-C-C-A (spot T2 of Table Va) and using the angular mobility shifts between successive oligonucleotide spots (Fig. 10A), the sequence that can be deduced for the 3'-terminal end is ...(G)-A-G-G-G-U-C-A-U-C-G-G-A-A-C-C-A.

A similar method was employed to obtain the sequence of the 5'-terminal region of the TYMV RNA 3'-fragment. The fragment was labeled with $[^{32}P]$ at the 5'-end and repurified by gel electrophoresis as in Fig. 14. The 5'-$[^{32}P]$-labeled material was recovered, partially digested with nuclease P_1 and the digest analyzed by two-dimensional homochromatography. Fig. 26 shows the results obtained. Although the partial heterogeneity at the 5'-end of this fragment gives rise to several satellite spots near the top of the autoradiogram, the sequence which can nevertheless be deduced (legend to Fig. 26) for the 5'-terminal region is ...U-A-A-G-U-U-C-U-C-G-A-U-C-U-U-U-A-A-A-A... . Based on the overlap of this sequence with the

Figure 25. Autoradiogram of a partial nuclease P_1 digest of 3'-[^{32}P]-labeled TYMV RNA. First dimension ionophoresis on cellulose acetate, pH 3.5, and second dimension, homochromatography in 10 mM KOH-strength "homomix" (Methods).

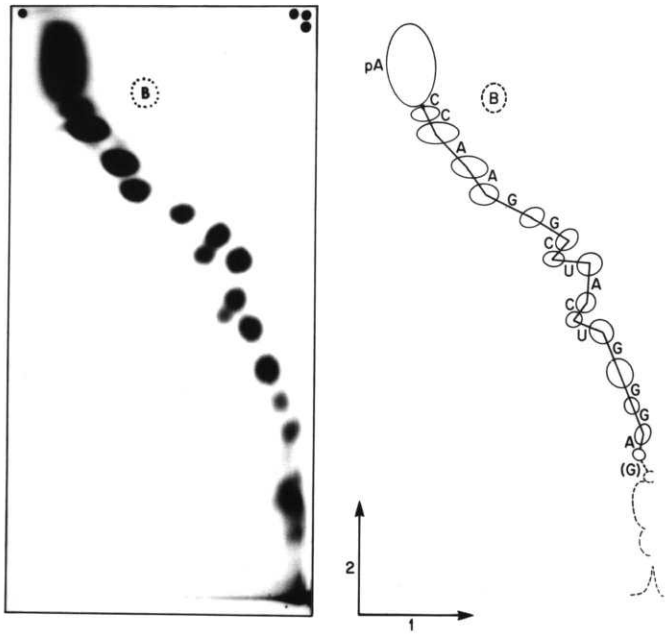
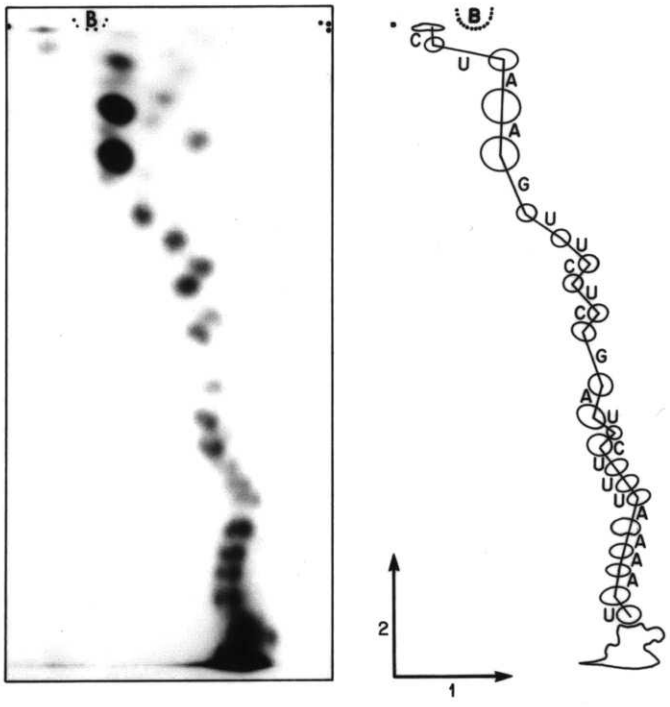

B
pA
C
C
A
A
G
G
C
U
A
C
U
G
G
G
A
(G)
B
2
1

Figure 26. Autoradiogram of a partial nuclease P_1 digest of 5'-[^{32}P]-labeled TYMV RNA 3'-fragment as analyzed by two-dimensional homochromatography (using 10 mM KOH-strength "homo-mix"). Note that the 5'-proximal [^{32}P]-labeled mono and di-nucleotides have been allowed to run off the top of the plate into the wick (the longer run increases resolution of material at the bottom of the plate). The "reading" of this sequence in the presence of numerous background spots was aided greatly by the recognition, within this pattern, of the known homochromatography patterns of the T_1-oligonucleotides (Table V) it contains.


B
B
2
1

oligonucleotides present in the complete T_1 RNase digest of the TYMV RNA 3'-fragment (Table Va), the 5'-terminal sequence of the TYMV RNA 3'-fragment is A-C-C-U-A-A-G-U-U-C-U-C-G-A-U-C-U-U-U-A-A-A-A-U-C-G.

Gel electrophoretic analysis of some preparations of the TYMV RNA 3'-fragment, 5'-terminally labeled with [^{32}P], showed the additional presence of a discrete band migrating slightly more slowly than the main band (Fig. 27). Analysis of a partial nuclease P_1 digest of this material is presented in Fig. 28. While this material also shows a few satellite spots and hence is not homogeneous at the 5'-end, it is nevertheless possible to conclude that this fragment contains the sequence ...U-U-C-C-A-C-C-U-A-A-G-U-U-C... .

The overlap of sequence between this and the 5'-end of the TYMV RNA 3'-fragment (Fig. 26) suggests that the sequence A-C-C-U-A-A-G.....A-A-A-A-U-C-G is preceded by U-U-C-C: ...U-U-C-C-A-C-C-U-A-A-G-U-U-C-U-C-G-A-U-C-U-U-U-A-A-A-A-U-C-G... .

A possible secondary structural model of interaction between this 5'-end sequence and the 3'-end sequence of the TYMV RNA 3'-fragment (Fig. 25) is discussed below (Discussion).

III. Further Applications of these Sequencing Techniques

A. The First Nucleotide Sequence of an Organelle tRNA: Euglena gracilis tRNA^{Phe}

Although the sequences of some 80 different tRNAs from a variety of sources are known, none of these represent

Figure 27. Autoradiogram of 5'-[^{32}P]-labeled TYMV RNA 3'-fragment subjected to electrophoresis in a 12% polyacrylamide gel. Two different preparations of the TYMV RNA 3'-fragment were run side by side. 1) gave two bands of approximately equal intensity, labeled A and B, and 2) gave the more typical single band (cf. Fig. 14) and serves here as a marker. Direction of electrophoresis was from top to bottom.

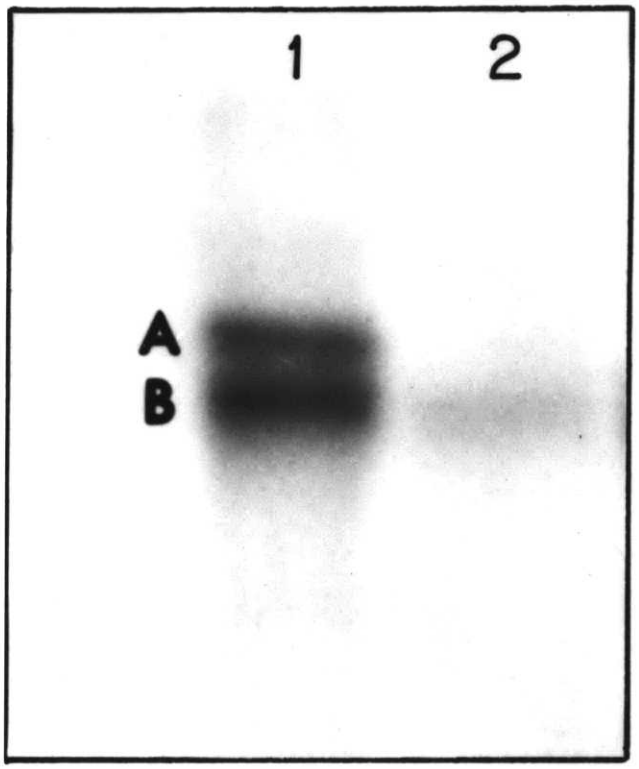
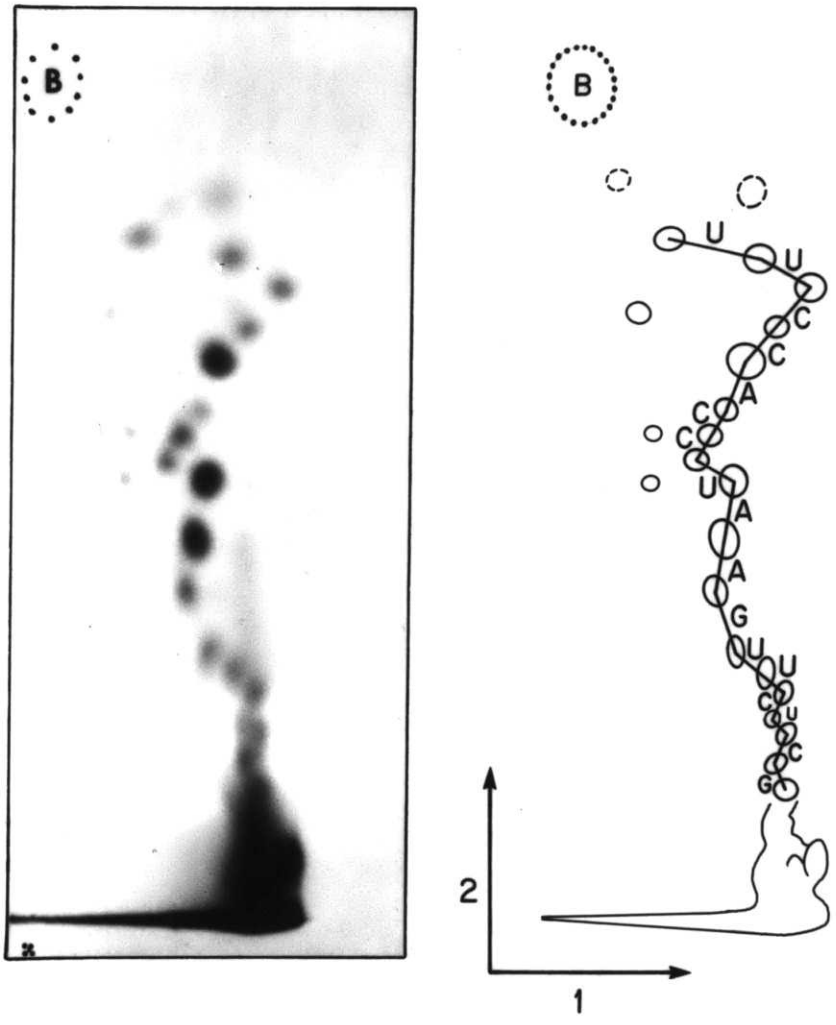

1
2
A
B

Figure 28. Autoradiogram of a partial nuclease P_1 digest of the slightly longer 5'-[^{32}P]-labeled TYMV RNA fragment isolated from band A of Fig. 27. Note that the mono- (pA 5'-end group) and dinucleotides (and probably several further oligonucleotides) have not been transferred from the first dimension to this plate. Because of the nature of "mobility shifts" in two-dimensional homochromatography (Fig. 10), the sequence at the 5'-end of this RNA fragment must be *pA-C or *pA-A, dinucleotides that run slower than xylene cyanole dye (B circled by dots) in the first dimension (ionophoresis on cellulose acetate at pH 3.5; vide ante). The exact size of the first oligonucleotide included in the "reading" proposed is not known; its position in the second dimension is consistent with a tetra-hexanucleotide.


B
B
U
U
C
C
A
C
C
U
A
A
G
U
U
C
C
G
C
G
2
1

tRNAs from eukaryotic organelles. The main reason for this has been the difficulty, until now, of obtaining either uniformly labeled [^{32}P]-tRNA or substantial quantities of non-radioactive tRNA from eukaryotic organelles in a homogeneous form for sequence studies. The purification and nucleoside composition of tRNA^{Phe} from chloroplasts of Euglena has been reported (155). The nucleotide sequence of this tRNA has now been determined, in collaboration with S. Chang and C. Brum, who worked in our laboratory (153,154); tRNA for sequence analysis was prepared by L. Hecker in the laboratory of W.E. Barnett. The methods used in this work followed, for the most part, the scheme presented in Fig. 12.

The nucleoside composition of chloroplastic tRNA^{Phe} was already known (155), so the application of [^{32}P]-end group labeling began with the preparation of fingerprints. This work was in collaboration with S. Chang, following procedures developed previously for the sequence analysis of the TYMV RNA 3'-fragment (vide ante) and the sequence analysis of eukaryotic initiator methionyl tRNAs (16,18,19).

Figure 29 shows the fingerprints obtained from complete T_1 RNase and pancreatic RNase digests of the chloroplastic tRNA^{Phe}. The 5'-[^{32}P]-labeled oligonucleotides were eluted and analyzed further by 5'-end-group analysis and by partial digestion with SVP and/or nuclease P_1, as described above (Section I). Table VI lists the sequences of these oligonucleotides. The elucidation of the sequences of the T_1 RNase

Figure 29. Fingerprints of 5'-[^{32}P]-labeled oligonucleotides obtained from (A) T_1 RNase digestion and (B) pancreatic RNase digestion of chloroplastic $tRNA^{Phe}$. B circled by dots is xylene cyanole dye marker.

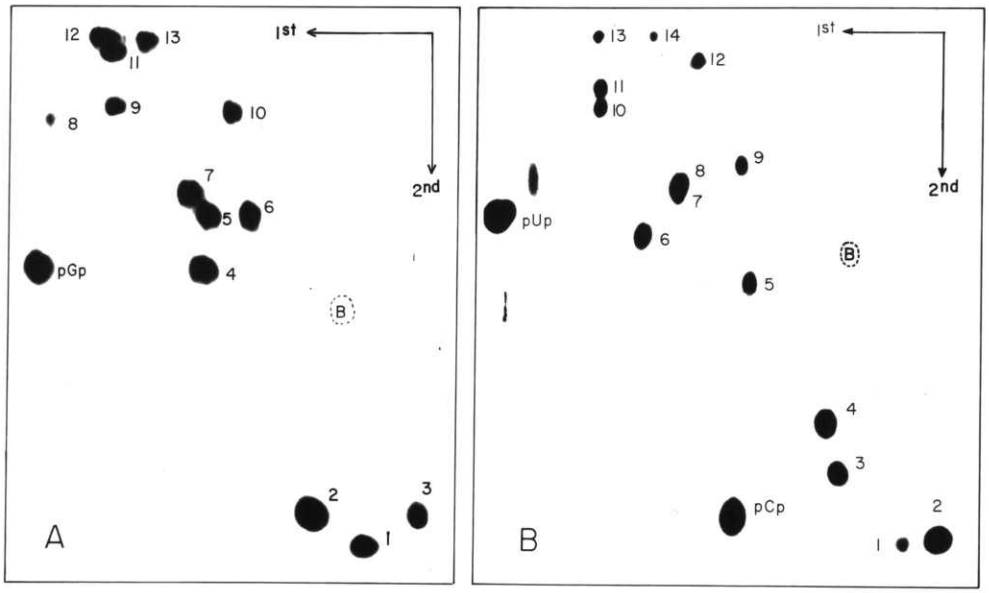

12
13
11
1st
2nd
8
9
10
7
5
6
pGp
4
B
2
3
1
A
13
14
12
11
10
8
9
7
pUp
6
5
4
3
pCp
2
1
B

Table VI. Oligonucleotide Sequences and Molar Yields Present in Fingerprints (Fig. 29 A and B) of T_1 RNase (A) and Pancreatic RNase (B) Digests of Euglena chloroplastic $tRNA^{Phe}$

---

[a] See Fig. 29

[b] Based on Cerenkov counting of fingerprint spots.

[c] Py* behaves like a uridine derivative in the two dimensional homochromatography system (Results I). It is probably a derivative of 4(abu)^{3}U, which became modified during the phenoxyacetylation of the tRNA during purification (155,158,159).

[d] Low molar ratio is due to the presence of modified nucleotides at the 5'-ends of these fragments, that are often phosphorylated at a slower rate by polynucleotide kinase.

[e] U* is either U or a modified U, probably D (dihydrouridine).

[f] A* is probably ms^{2}i^{6}A, known to be present in chloroplastic $tRNA^{Phe}$ (155).

Spot No.[a]	Sequence	Molar Ratio[b]
T1	C-G	1.2
T2	A-G	2.4
T3	C-A-C-C-A	0.92
T4	C-U-G	1.1
T5	A-C-U-G	1.0
T6	C-U-C-A-G	1.0
T7	A-U-A-G	1.0
T8	D-U-G	0.24[d]
T9	Gm-U*-A-G[e]	0.53
T10	U-m^7G-Py*-C-A-C-C-A-G[c]	0.68
T11	U-U-C-C-U-A-G	0.94
T12	T-Ψ-C-A-A-A-U-C-U-G	1.0
T13	A-A-A*-A-Ψ-C-C-U-U-G[f]	0.61
A1	m^7G-Py*[c]	0.2[d]
A2	A-C	2.0
A3	G-C	1.0
A4	A-G-C	2.0
A5	A-A-A-U	1.0
A6	G-U	1.0
A7	A-G-D	1.0
A8	A-G-T	1.0
A9	A-G-A-G-C	1.0
A10	G-Gm-U*[e]	0.75
A11	G-G-U	1.0
A12	G-A-A-A*-A-Ψ[f]	0.63
A13	G-G-G-A-U	1.0
A14	G-G-A-G-G-A-C	1.0

and pancreatic RNase oligonucleotides arising from the anticodon loop region of the tRNA (T13 and A12), in particular, relied on a combination of analyses with SVP and nuclease P_1 (data not shown), as described above for the analysis of the anticodon sequence of Neurospora crassa $tRNA_i^{Met}$ (Fig. 11).

Determination of the total sequence of the tRNA required, in addition, sequence information for the 3'- and 5'-ends of the molecule, and sequence information for larger internal oligonucleotides. Sequence information for the 3'-end of the tRNA was obtained by specific chemical cleavage of the tRNA at the phosphodiester bond adjacent to m^7G (16,156); this was the work of C. Brum. The smaller 3'-fragment and larger 5'-fragment of the tRNA were separated by polyacrylamide gel electrophoresis in 7 M urea (Methods), and recovered from the gel by extraction (Methods). The sequence of the 3'-terminal, 30-nucleotide-long fragment was established by complete digestion with T_1 RNase and complete digestion with pancreatic RNase, followed by 5'-end labeling of the oligonucleotides with $[^{32}P]$ and fingerprinting. The sequence of the 5'-end of the tRNA was determined by partial nuclease P_1 digestion of 5'-$[^{32}P]$-labeled intact tRNA, and analysis by two-dimensional homochromatography (data not shown). Several internal fragments of the tRNA were obtained by partial digestion, under mild conditions, with T_1 RNase (157). The larger oligonucleotide fragments produced were treated with calf-intestinal alkaline phosphatase to remove 3'-terminal phosphomonoester groups

(Methods), and were labeled with [^{32}P] at the 5'-end; the labeled fragments were then separated by two-dimensional gel electrophoresis (Fig. 30). The 5'-[^{32}P]-labeled fragments were recovered by electrophoretic elution (Methods) and analyzed either by partial digestion with nuclease P_1 (followed by two-dimensional homochromatography) or by further complete digestion with T_1 RNase followed by [^{32}P]-labeling and fingerprinting of the oligonucleotides present in such digests (16). The analysis and sequence of selected oligonucleotide spots (Fig. 30) that were characterized is summarized in Table VII.

The use of all these analyses to order the fingerprint oligonucleotides into a unique total sequence for chloroplastic $tRNA^{Phe}$ is summarized in Fig. 31. The sequence at the top is that of the tRNA in linear form. The sequence shown in (a) was deduced from a partial nuclease P_1 digestion of 5'-[^{32}P]-labeled tRNA. Fragments (b) and (e) are larger oligonucleotides present in partial T_1 RNase digests (Table VII). The sequence shown in (b) was obtained by partial digestion with nuclease P_1, whereas that shown in (e) was established by analysis of the oligonucleotides present in complete T_1 RNase digests of (e). Fragment (d) is the 3'-terminal fragment obtained by chemical cleavage of the tRNA adjacent to the m^7G residue, and its sequence was established as described above. Oligonucleotides (d), (f), and (h) are among those present in complete T_1 RNase and pancreatic RNase

Figure 30. Autoradiogram of two-dimensional polyacrylamide gel electrophoresis of 5'-[^{32}P]-labeled partial T_1 RNase digest of chloroplastic $tRNA^{Phe}$ (details in text). XCB, xylene cyanole dye; BøB, bromophenol blue dye. The first dimension is a 10% gel run at pH 3.5, and the second dimension a 20% gel run at pH 8 (Methods).

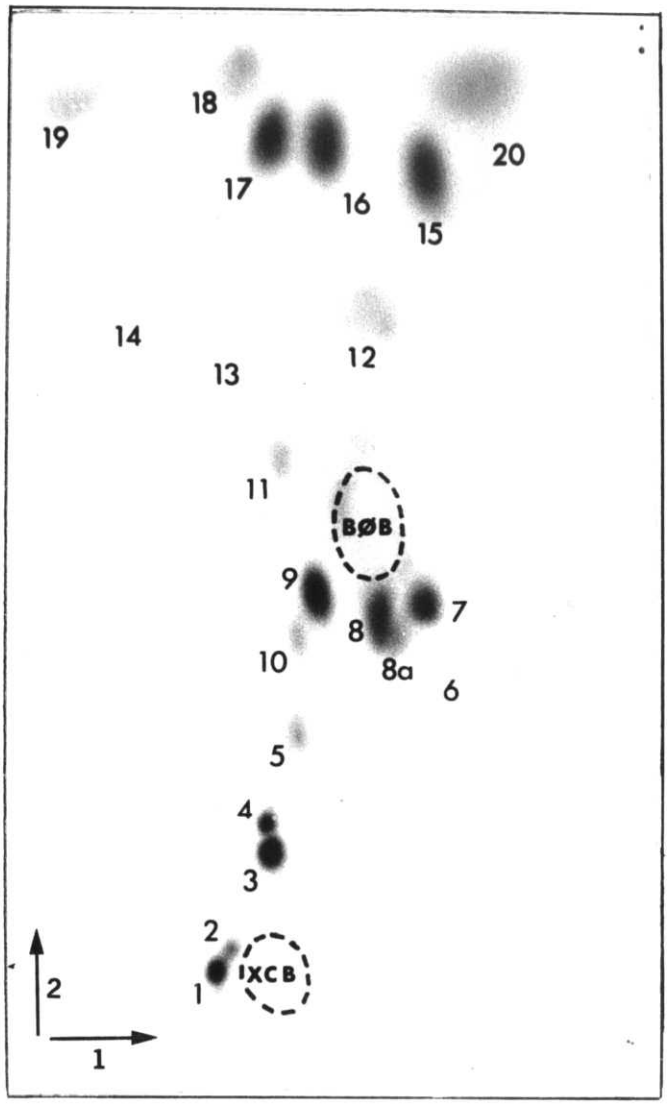

19
18
17
16
15
20
14
13
12
11
BØB
9
8
7
10
8a
6
5
4
3
2
1
IXCB
2
1

Table VII. 5'-end Groups and Sequences of Some 5'-[^{32}P]-labeled RNA Fragments Recovered after Partial T_1 RNase Digestion of Chloroplast $tRNA^{Phe}$ (Fig. 30).

a See Fig. 30.

b Determined by digestion with T_2 RNase, followed by one-dimensional TLC (Methods; cf. Fig. 6); n.d., not determined.

c Sequential nucleotide numbering, Fig. 31.

d Comp. T_1 refers to complete digestion with T_1 RNase, followed by 5'-end group labeling of the resulting oligonucleotides and fingerprinting. Part. P_1 refers to partial digestion with nuclease P_1, followed by two-dimensional homochromatography.

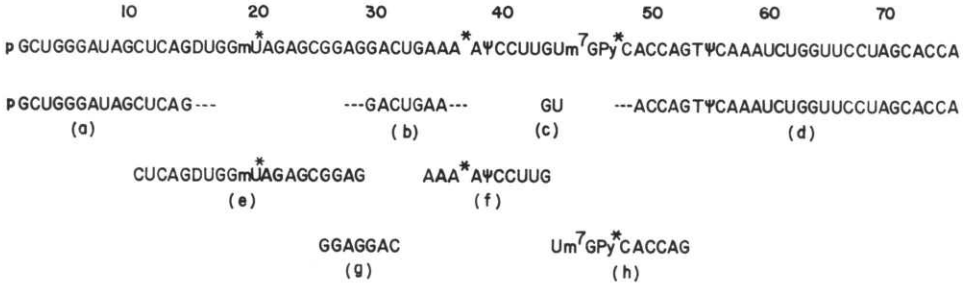
e Letters in parentheses refer to fragment designations in Fig. 31.

f The 5'- and 3'-portions of this nucleotide sequence could not be "read" unambiguously from the two-dimensional homochromatogram (data not shown).

Table VII. 5'-end Groups and Sequences of Some 5'-[^{32}P]-labeled RNA Fragments Recovered after Partial T_1 RNase Digestion of Chloroplast $tRNA^{Phe}$ (Fig. 30).

Frag. No.[a]	5'-End Group[b]	Sequence[c]	Method of Analysis[d]
1	A	35→76	Comp. T_1
2	A	35→ ?	Comp. T_1
3	U	45→ ?	Comp. T_1
4	U	45→ ?	Part. P_1
6 (b)	A	...G-A-C-U-G-A-A[f]...	Part. P_1
8 (e)	n. d.	11→29	Comp. T_1

Figure 31. Sequence of chloroplastic $tRNA^{Phe}$ in linear form (top) and large oligonucleotides necessary for establishment of the sequence. Fragments (a) through (h) are explained in the text. Dashed lines indicate that the fragments concerned contained additional nucleotides, the sequence of which was not completely deduced in the particular experiment described.

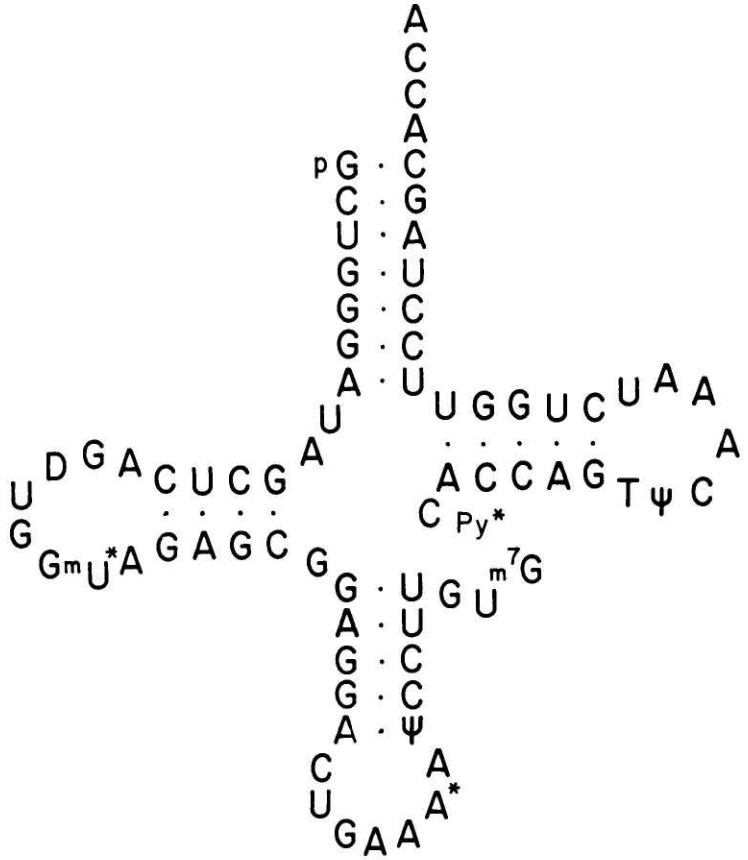

10 20 30 40 50 60 70
p GCUGGGAUAGCUCAGDUGGmU*AGAGCGGAGGACUGAAA*AΨCCUUGUm7GPy*CACCAGTΨCAAAUCUGGUUCCUAGCACCA
pGCUGGGAUAGCUCAG---
(a)
---GACUGAA---
(b)
GU
(c)
---ACCAGTΨCAAAUCUGGUUCCUAGCACCA
(d)
CUCAGDUGGmU*AGAGCGGAG
(e)
AAA*AΨCCUUG
(f)
GGAGGAC
(g)
Um7GPy*CACCAG
(h)

digests of the tRNA (Table VI) and are shown here only to indicate the overlaps necessary for the derivation of the total sequence of the chloroplastic tRNA.

Fig. 32 shows the sequence of Euglena chloroplastic $tRNA^{Phe}$ written in the cloverleaf form (Fig. 1). U*, A*, and Py* are three modified nucleosides which have not been conclusively identified during this sequence work. U* is either U or D. Based on the behavior of oligonucleotides containing A* and on earlier work on the nucleoside composition analysis of the tRNA (155), A* is most probably ms^2i^6A (2-methylthio-N^6-isopentenyladenosine) and Py* a phenoxyacetylated derivative of $4(abu)^3U$ (N^3-(4(2-amino)butyryl)uridine) (158,159).

In the context of the issues raised in this thesis, the elucidation of the total nucleotide sequence of chloroplastic $tRNA^{Phe}$ provides an excellent example of the application of [^{32}P]-end-group-labeling to RNA sequence analysis. Organellar tRNAs are difficult to purify in substantial quantities, and cannot be radiolabeled to high specific activity in vivo. Our analysis of chloroplastic $tRNA^{Phe}$ represents the first complete nucleotide sequence to be determined for any organellar tRNA, and illustrates both the sensitivity and efficiency of the [^{32}P]-post-labeling sequencing methods described in Section I; most of the work described above (except the m^7G cleavage reaction) was completed with a total of less than 75 μg of RNA in a time span of four months.

Figure 32. Sequence of Euglena chloroplast $tRNA^{Phe}$ in clover-leaf form. Modified nucleosides are discussed in text.


A
C
C
A
pG · C
C · G
U · A
G · U
G · C
G · C
A · U
U
U G G U C U A A
A
D G A C U C G
A C C A G T ψ C
U
G
C Py*
Gm U* A G A G C G
G · U G U m7G
A · U
G · C
G · C
A · ψ
C A
U A*
G A A

B. Analysis of Sites of Chemical Modification of 5S RNA in Bacterial Ribosomes

It has been shown in bacteria, that the presence of 5S RNA in the 50S ribosomal subunit is essential for ribosomal activity (160). Moreover, it has been found that the tRNA fragment T-Ψ-C-G specifically interacts with 5S RNA (161). It was also found that T-Ψ-C-G inhibits enzymatic binding of aminoacyl tRNA to ribosomes and stimulates magic spot formation (162). Because the sequence G-U-U-C-purine, and its derivatives, such as G-T-Ψ-C-purine, are common to loop IV of all tRNAs active in peptide chain elongation (Fig. 1), this and other evidence suggest that 5S RNA interacts with this sequence of aminoacyl tRNA in the ribosomal A-site. Furthermore, the existence of a G-A-A-C sequence, complementary to T-Ψ-C, in approximately the same position in 5S RNA isolated from a variety of bacteria (11) suggests that this sequence may be the locus of such interaction, which could occur by Watson-Crick base-pairing.

When 5S RNA is treated in situ, on the 70S ribosome, with monoperphthalic acid, a selective oxidation of two adenine residues occurs (161). 5S RNA so treated loses its ability to bind T-Ψ-C-G, and 50S ribosomal subunits reconstituted with treated 5S RNA have greatly reduced activity (161). Because of the proposed role of the G-A-A-C sequence in the function of 5S RNA (i.e., Watson-Crick base-pairing with T-Ψ-C), it was suggested that the sites of adenosine

modification by monopherphthalic acid in 5S RNA may be within this sequence (161).

In collaboration with V.A. Erdmann of the Max-Planck-Institut für Molekuläre Genetik (Berlin-Dahlem, W. Germany), we decided to examine the sites of *in situ* oxidation of 5S RNA by monoperphthalic acid through sequence analysis. The analytical methods involved pancreatic RNAase digestion of 5'-[^{32}P]-labeling of the resulting oligonucleotides (which include the oligonucleotide G-A-A-C), and fingerprinting; oligonucleotide sequences were determined by partial SVP digestion followed by two-dimensional homochromatography.

5S RNA treated *in situ*, in the ribosome, with monoperphthalic acid as described (161) was prepared by V.A. Erdmann; samples were treated for 5, 15 and 45 min, respectively. Fig. 33 shows the fingerprints obtained from complete pancreatic RNase digests of these monoperphthalic acid-treated 5S RNA samples, and from an untreated 5S RNA control. It can be seen (Fig. 33, arrows) that two major new oligonucleotide spots appear with increasing time of exposure of the RNA to monoperphthalic acid, the spots numbered 13* and 18*, as well as a third product, numbered 20. One additional new product (Fig. 33, arrows) was present only in the 45 min treated RNA (spot 21); it was not further studied.

The sequences and molar yields of all the oligonucleotides in the four fingerprints (Fig. 33) are indicated in Table VIII. Identification of sequences was based mostly on

Figure 33. Fingerprints obtained from pancreatic RNase digestions of 5S RNA which had been treated with monoperphthalic acid for A) 0 min; B) 5 min, C) 15 min and D) 45 min. New spots appearing during this time course (13*, 18*, 20, 21) are indicated by arrows. Numbering of parent spots is according to Brownlee (129). Some spots which appeared only in single fingerprints were assumed to be artifactual (i.e., caused by partial degradation of the substrate before initiation of the fingerprinting analysis) and were not numbered. Such decisions must still be considered preliminary at this time. Spot X is a radioactive artifact seen often in pancreatic RNase fingerprints prepared by our in vitro labeling procedures (M. Simsek, personal communication). Spot Y is in a position consistent with the identification *pA-A-C, although this oligonucleotide should not be present in a pancreatic RNase digest of E. coli 5S RNA (129). Because of its variable yield (Table VIII), spot Y may be due to artifactual nuclease degradation.

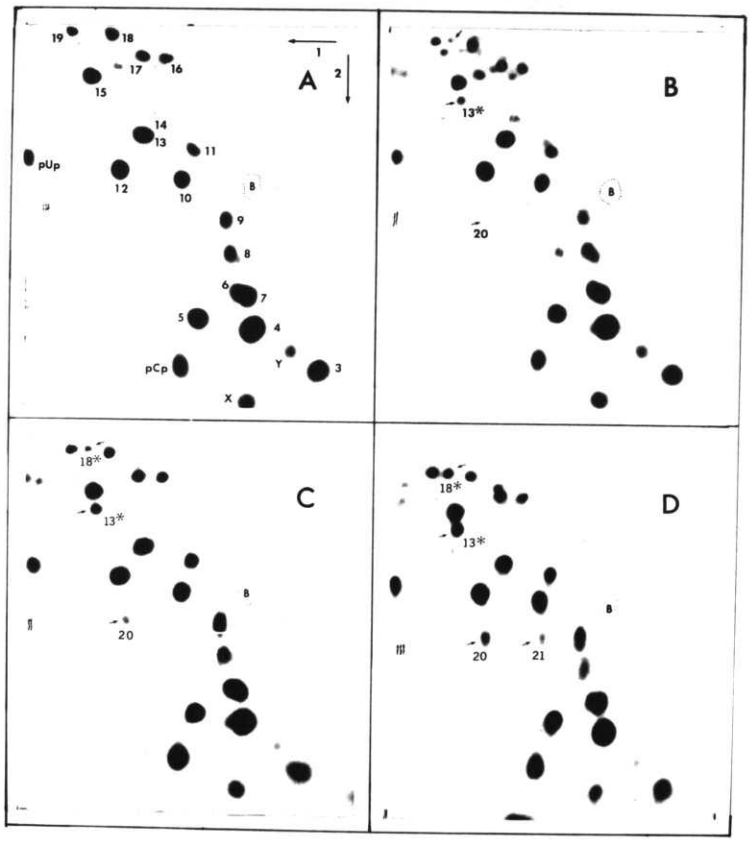

A
B
C
D
1
2
19
18
17
16
15
14
13
11
pUp
12
10
9
8
6
7
5
4
pCp
Y
3
X
18*
13*
20
21

Table VIII. Oligonucleotide Sequences and Molar Yields Present in Fingerprints (Fig. 33, A-D) of Pancreatic RNase Digests of E. coli 5S RNA Treated for A) 0 min, B) 5 min, C) 15 min, and D) 45 min with Monoperphthalic Acid (161; see text)

---

[a]See Fig. 33 A-D.

[b]Based on Cerenkov counting of excised fingerprint spots. The molar yields of spots 13 and 14 were calculated from the ratio of A to G end groups in material from the mixed spot 13 + 14 (Fig. 33 A-D).

[c]Although the sequence of this oligonucleotide was not determined in this work, the indicated sequence was inferred from its position in the fingerprint (Fig. 33D; see text).

Spot No.[a]	Sequence	Molar Ratio[b]				
		A	B	C	D	expected (ref. 129)
X	--	0.92	1.5	0.98	1.0	
Y	--	n.d.	0.6	0.2	0.26	
3	A-C	2.0	2.4	1.4	1.8	1
4	G-C	6.4	7.2	9.1	9.2	7
5	A-U	2.8	2.9	2.9	3.2	3
6+7	1G-A-C + 2A-G-C	3.6	3.8	3.5	3.1	1+2
8	G-A-A-C	0.95	1.6	1.0	0.99	1
9	G-A-A-A-C	0.89	0.9	1.6	1.6	1
10	G-G-C	2.1	1.8	2.3	2.7	2
11	A-G-G-C	0.89	0.92	0.93	1.1	1
12	G-U	3.6	4.8	3.4	4.0	2.5
13	G-A-U[b]	1.3	1.4	0.74	0.31	1
14	A-G-U[b]	1.6	1.8	1.7	1.9	1
15	G-G-U	3.0	2.9	3.5	3.5	3
16	A-G-G-G-A-A-C	0.65	0.68	0.61	0.55	1
17	A-G-A-A-G-U	0.77	0.96	1.1	0.96	1
18	G-A-G-A-G-U	0.87	0.94	0.79	0.51	1
19	G-G-G-G-U	0.47	0.38	0.60	0.74	1
13*	G-A*-U	--	0.32	0.52	1.2	
18*	G-A*-G-A-G-U	--	--	0.15	0.50	
20	(A*-U)[c]			0.22	0.61	

5'-end group analysis, and comparison with the fingerprint patterns of Brownlee and Sanger (129); several identifications were further corroborated by partial SVP digestion, followed by two-dimensional homochromatography (data not shown). The two oligonucleotides whose yield decreases with time of exposure to monoperphthalic acid (Table VIII) are 13, which goes from 1.2 mole to 0.18 mole, and 18, which goes from 0.87 mole to 0.51 mole. The oligonucleotides that appear during the time course of exposure to monoperphthalic acid (Fig. 33, Table VIII) may be seen to folow the reverse kinetics, 13* being inversely proportional to 13, and 18* (and also 20) being inversely proportional to 18.

The analysis of the sequence of 13* and 18* by two-dimensional homochromatography is shown in Fig. 34; Fig. 35 shows, for comparison, the pattern obtained, by this method, from unmodified oligonucleotide 18 (Fig. 33; Table VIII). The mobility shift analysis (Section I) of the sequence of 13* and 18* (Fig. 34) was routine, except for the first mobility shift, from mono- to dinucleotide, in both 13* (Fig. 34A) and 18* (Fig. 34B), which was found to be different than that of any of the four major nucleotides (Fig. 10A). This mobility shift, labeled A* in Fig. 34, was essentially identical to a G-shift in the first dimension, cellulose acetate ionophoresis at pH 3.5 (Fig. 10A), but resembled a small pyrimidine shift (see Section I, above) in the second, homochromatography dimension. The expected product

Figure 34. Analysis of the sequences of the oligonucleotides represented by spots 13* and 18* in the fingerprints of Fig. 33. Partial SVP digests of the 5'-[^{32}P]-labeled oligonucleotides are analyzed by two-dimensional homochromatography. Designation of the A* mobility shift is discussed in the text.

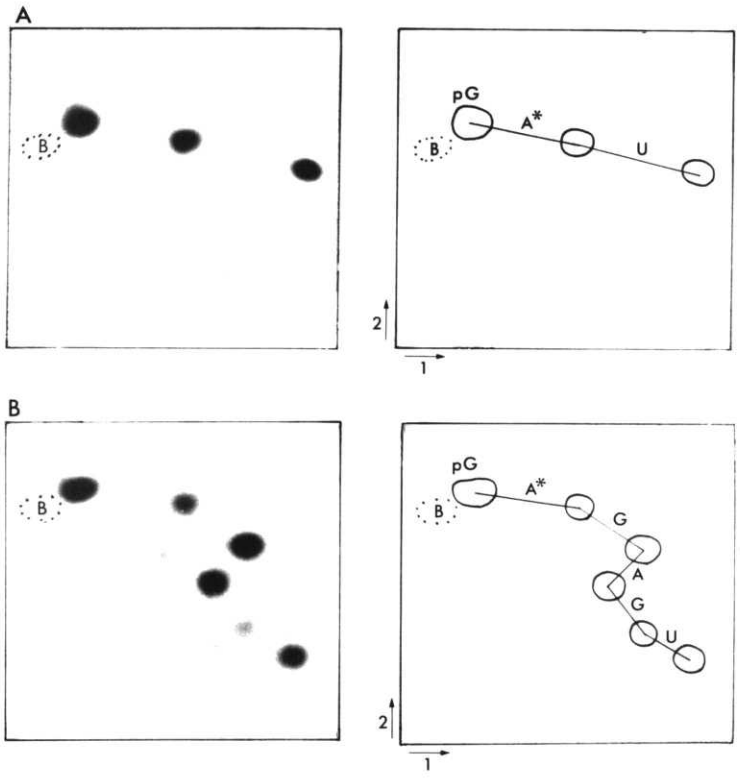
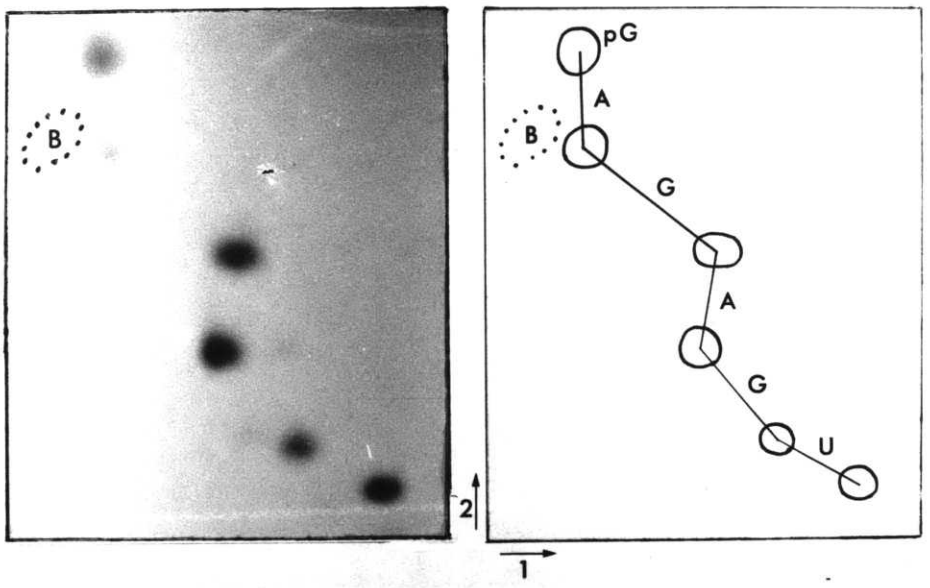

A
B
pG
A*
U
2
1
B
B
pG
A*
G
A
G
U
2
1

Figure 35. Autoradiogram of analysis of *pG-A-G-A-G-U (Fig. 33, spot 18) by two-dimensional homochromatography of a partial SVP digestion of the 5'-[^{32}P]-labeled oligonucleotide.


B
pG
A
B
G
A
G
U
2
1

of monoperphthalic acid oxidation, under the conditions used in this experiment (these conditions favor reaction with A residues that are not base-paired (163)), is indicated in Fig. 36. This product, adenosine-N-1-oxide, may have an altered acid pK (cf. Table IV), and a "G-like" behavior upon ionophoresis at pH 3.5 would therefore not be surprising[1].

Overall, the evidence suggests strongly that 13* is G-A*-U[2] and 18* is G-A*-G-A-G-U[2]. First, it has been previously reported (161) that application of the conditions used in this experiment to monoperphthalic acid modification of 5S RNA in situ in the ribosome results specifically in oxidation of two A residues. Second, our data show that the kinetics of the appearance of 13* and 18* matches the kinetics of the loss of 13 and 18, respectively (Table VIII). Third, the first-dimensional mobility, in fingerprints, of 13*, which runs with G-G-U (Fig. 33, spot 15) and of 18*, which runs faster than 18 (Fig. 33) can be explained as resulting from the "G-like" behavior of adenosine-N-1-oxide at pH 3.5. Finally, the sequence analysis of 13* and 18* by partial SVP digestion followed by two-dimensional homochromatography (Fig. 34) is entirely consistent with the sequence assignments G-A*-U and G-A*-G-A-G-U, respectively. The identity of oligonucleotide 20 (Fig. 33) was not investigated

---

[1] The behavior of A* on DEAE-paper ionophoresis in 7% formic acid (i.e., the second dimension of a standard fingerprint), or on homochromatography cannot be predicted on a priori grounds.

[2] A* refers to a modified A-residue, in this case adenosine-1-oxide.

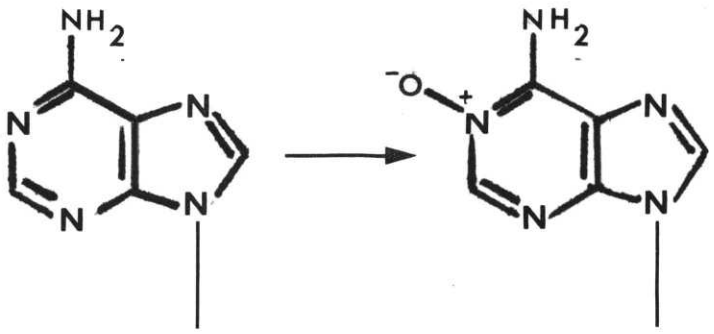
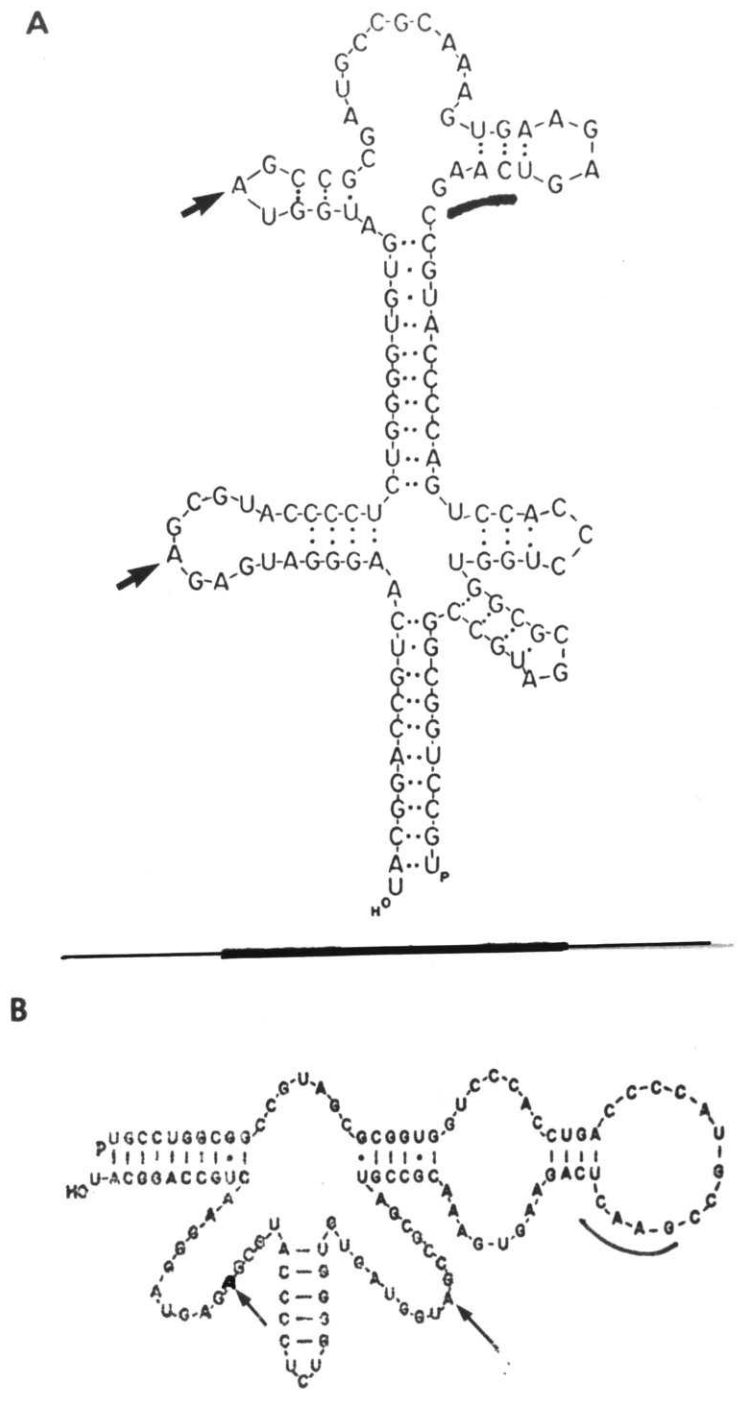


Figure 36. Oxidation of adenine residues in a polynucleotide chain by monoperphthalic acid (163).

in this work; however, its first dimensional fingerprint mobility (Fig. 33), which is very similar to that of G-U (Fig. 33, spot 12), suggests that it is probably A*-U.

Based on the above analysis two of the major sites of modification in 5S RNA that result from treatment, in situ on 70S ribosomes, with monoperphthalic acid may be identified with residues 73 and 99 in the primary sequence of the molecule (Fig. 37). Since there are three A-U sequences in 5S RNA (99; Fig. 37), the locus or loci of the A*-U modification cannot yet be determined from this work. In several proposed secondary structural models for 5S RNA (Fig. 37), the two identified sites are in single-stranded regions. However, as these modifications were generated by treatment of the RNA in situ, in the ribosome, the significance of our results lies in the implication that these two sites are also "exposed" on the ribosome. In this context, it is especially interesting that only one of the two A residues of oligonucleotide fragment 18 (Fig. 33; Table VIII) is subject to modification. Unfortunately, the relationship between monoperphthalic acid oxidation of two exposed A-sites on 5S RNA and loss of the ability of this RNA to interact with T-Ψ-C-G (161) is not yet clear. If the G-A-A-C sequence of 5S RNA (residues 44-47 in Fig. 37) is, in fact, the actual locus of interaction with T-Ψ-C-G by Watson-Crick base-pairing (an unproven hypothesis), our results imply that tertiary structural interactions, either in the 5S RNA itself or in its situation

Figure 37. Secondary structural models for E. coli 5S RNA, indicating (arrows) two of the loci of modification by monoperphthalic acid (see text). Model A is from ref. 164 and model B is from ref. 165.


A
B

within the ribosome, must also play a major role in this interaction.

Our experiments with chemically modified 5S RNA also point out the versatility both of the use of 5'-end group labeling for comparative fingerprinting, and of the use of two-dimensional homochromatography for sequence analysis. The use of 5'-end group post-labeling, in this instance, rather than *in vivo* labeling with [^{32}P], also served to facilitate the collaborative aspect of the project; samples of 5S RNA prepared in Germany could be stored and analyzed at a later date, without concern over radioautolytic damage or the short half-life of [^{32}P].

DISCUSSION

In an attempt to account for the tRNA-like behavior of the 3'-terminal region of TYMV RNA, we have investigated the primary nucleotide sequence of a circa 112 nucleotide 3'-terminal TYMV RNA fragment. The sequence information we have obtained reveals some interesting characteristics of this RNA fragment. In addition, by comparing our data with that of L. Hirth and collaborators (104), a tentative sequence for the 115 3'-nucleotides of TYMV RNA may be proposed and a possible 'cloverleaf' secondary structural model examined. Finally, our efforts to devise means of efficiently deriving sequence information from non-radioactive RNA by *in vitro* labeling with [^{32}P] have resulted in several useful methodological advances; our sequencing strategies are examined with reference to other post-labeling sequencing methods currently in use.

I. The 3'-End of TYMV RNA

A. Characteristics of the tRNA-like 3'-Terminal Region of TYMV RNA

Our results show that unlike tRNA precursors in *E. coli* (167) and those of various eukaryotic organisms and tissue culture cells (168), the tRNA-like region of TYMV virion RNA does not contain any modified nucleosides. The TYMV RNA 3'-fragment is thus an example of an RNA species that can be enzymatically aminoacylated and is, in fact, purified on the

basis of aminoacylation (see Methods), but contains no modified nucleosides. This is clear proof that E. coli valyl-tRNA synthetase (ValRS), the enzyme used in the preparation of the TYMV RNA 3'-fragment, does not require the presence of modified nucleosides for cognate tRNA recognition. The same may also be true for ValRS of yeast, rat liver and plants, although, as only crude preparations of these enzymes have been tried with TYMV RNA (65,75), it is possible that some nucleoside modifications may be introduced (prior to aminoacylation) by tRNA modification enzymes present in the aminoacylation reaction mixtures.

Recently, it has been reported that TMV RNA can be aminoacylated by purified histidyl-tRNA synthetase (HisRS) from rat liver (84). Since the partial sequence reported for the 3'-71 nucleotides of TMV RNA does not contain any modified nucleosides (95), it is likely that TMV RNA, too, can be enzymatically aminoacylated in the absence of nucleoside modification. One interesting finding in the case of TMV RNA aminoacylation is that the enzymatic reaction is strongly inhibited by KCl (84); at 20 mM KCl, the aminoacylation of TMV RNA was found to be inhibited to 50%, whereas the reaction with rat liver tRNA was stimulated by KCl, the optimal concentration being 140 mM. This was interpreted (84) as suggesting that the binding of TMV RNA by rat liver HisRS is relatively weak, and may be readily dissociated by increasing salt concentrations. The aminoacylation of TYMV

RNA by E. coli ValRS has also been reported (75) to be inhibited by KCl (40 mM), although its aminoacylation by plant ValRS is not inhibited at the same salt concentration.

The aminoacylation of TYMV RNA and TMV RNA in the absence of nucleoside modification is consistent with the finding that E. coli $tRNA_1^{Val}$ in which nearly all the uridine and modified uridine residues have been substituted (in vivo) by the uridine analogue 5-fluorouridine is aminoacylated normally in vitro and has approximately the same Km for E. coli ValRS as normal $tRNA^{Val}$ (169), results which imply that the modified nucleosides that are uridine derivatives (such as T, Ψ, D, etc.), at least, are not essential for $tRNA^{Val}$ aminoacylation by E. coli ValRS.

Fluorouridine-substituted E. coli $tRNA_1^{Val}$ also appears to function normally in in vitro protein synthesis (170). In the case of TYMV RNA, on the other hand, it is not clear whether the 3'-end region can function in in vitro protein synthesis without nucleoside modification. Haenni, et al. (71), found that TYMV valyl-RNA could donate valine to polypeptides in an E. coli cell-free protein synthesis system only after a 10-15 min lag time (valyl-$tRNA^{Val}$ exhibited no lag). This result was later confirmed (70), and it was shown further that "RNase P"-treated TYMV valyl-RNA demonstrated an identical lag period prior to valine transfer. It must therefore be suspected that simple nucleolytic "processing" may not be the rate limiting step required for

ribosomal recognition of TYMV RNA, i.e., that nucleoside modification may also be involved. Alternatively, it is possible that TYMV valyl-RNA may not, in this system, donate valine to polypeptides at all, but may rather only transfer its valine moiety to endogenous $tRNA^{Val}$ by an enzymatically mediated transacylation reaction. The insensitivity of valine transfer by TYMV valyl-RNA to the valine aminoacylation inhibitor, valinol-AMP, has been used to argue that free valine cannot be an intermediate in the transfer of valine from TYMV valyl RNA to polypeptides (71). However, an enzymatically mediated transacylation that does not involve free valine cannot be excluded; if this reaction were rate-limiting, it could account for the observed lag time. Clearly, the function of TYMV RNA, or of the TYMV RNA 3'-terminal fragment, in protein synthesis requires furter investigation.

An interesting finding of our sequencing work is that the TYMV RNA 3'-fragment contains no G-U-U-C-purine sequence, as can be seen in Table V. This sequence feature has heretofore been found to be present in loop IV of all tRNAs active in polypeptide chain elongation (3), as indicated in Fig. 1. It may also be present, in unmodified form, in the 3'-terminal region of TMV RNA (95); TMV RNA can, in fact, act as substrate for the ribothymidine methylase of E. coli (78). The 3'-terminal sequence of the RNA of the Tymovirus, EMV, on the other hand, is lacking in this sequence (97). Because of the implication of this sequence feature in tRNA

interaction with ribosomes, via Watson-Crick base-pairing with a complementary sequence in 5S RNA (vide ante; ref. 161), it would be particularly important, in view of our findings, to re-examine the function of TYMV valyl-RNA in in vitro protein synthesis.

The use of nuclease P_1 on 3'-end labeled intact TYMV RNA and on the 5'-end labeled TYMV RNA 3'-fragment has enabled us to derive the sequence of longer stretches of nucleotides at the 3'- and 5'-ends of the TYMV RNA fragment. Based on these sequences, it is possible to examine whether the TYMV RNA fragment contains regions of complementarity between its 3'- and 5'-ends, as would be expected of a tRNA. Fig. 38 illustrates an example of a possible base-pairing scheme between the 3'- and 5'-ends of the fragment. Based on the thermodynamic considerations for RNA secondary structural stability of Tinoco et al. (171), the single looped-out nucleotide indicated in the figure is an acceptable perturbation in an 8-nucleotide long Watson-Crick base-paired helical region.

The nucleotide sequence at the 5'-end of the TYMV RNA fragment also suggests that this part of the fragment includes the end of the TYMV coat protein gene. From the data presented above, we have proposed that the sequence at the 5'-end of the fragment is ...U-U-C-C-A-C-C-U-A-A-G-U-U... in which UAA, a potential Ochre termination codon, is preceded by ACC, a threonine codon, and UCC, a serine codon. The

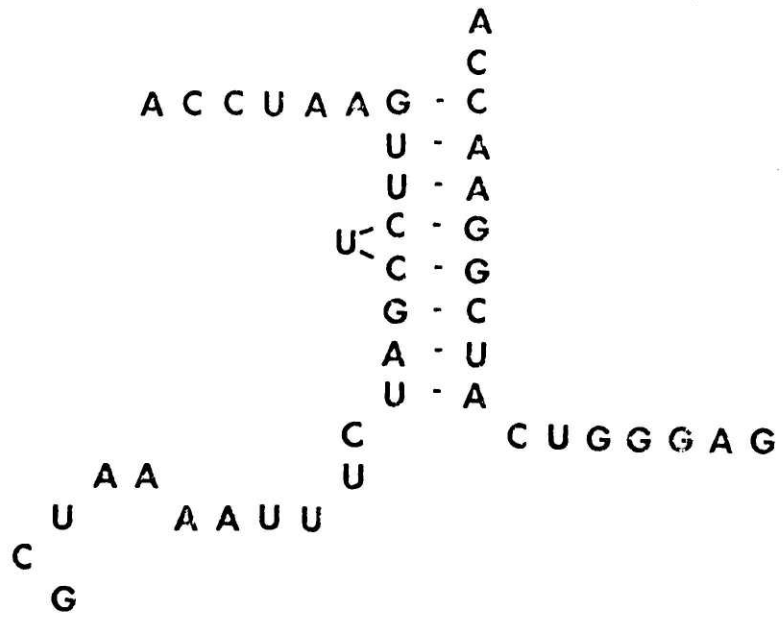


Figure 38. Possible base-pairing between the 3'- and 5'-end sequences of the TYMV RNA 3'-fragment.

reported primary amino acid sequence of the TYMV coat protein does, in fact, end in Ser-Thr at its carboxy terminus (172). It is therefore possible that the region of TYMV RNA which has tRNA-like properties follows immediately after the termination codon of the coat protein gene. Strong support for this hypothesis is provided by the independent sequencing work of L. Hirth and collaborators (104; vide infra).

B. Complete Nucleotide Sequence of the 3'-Terminal Region of TYMV RNA

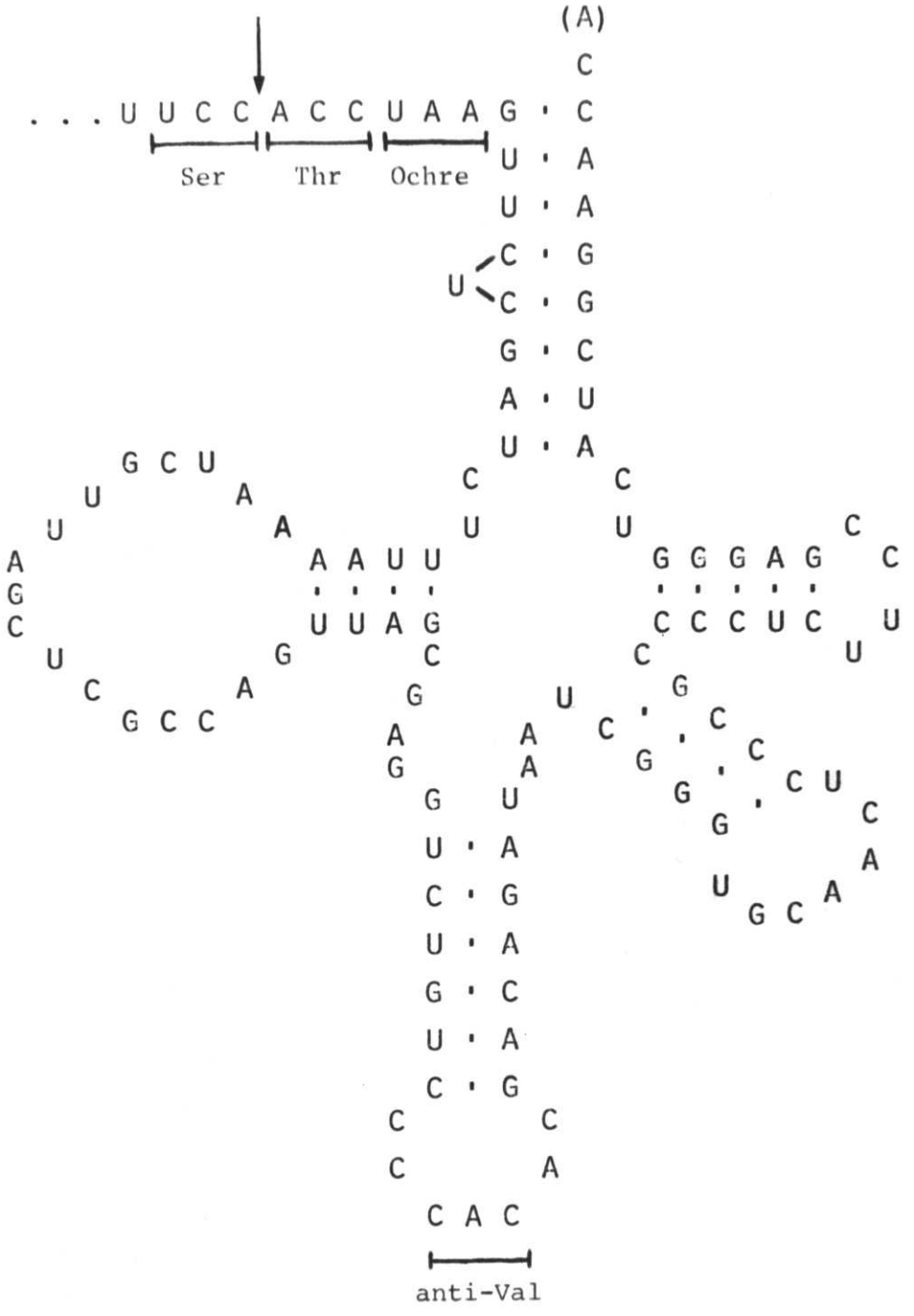
During the course of our investigation of the primary nucleotide sequence of the TYMV RNA fragment, we were apprised of similar work in progress in the laboratory of L. Hirth at the University of Strasbourg, France. We have since compared our respective results and found them to be in substantial agreement (104). Working with RNA labeled uniformly in vivo with [^{32}P], Hirth and collaborators have fingerprinted many overlapping 3'-end-specific RNA fragments generated by mild partial digestions of intact TYMV RNA with T_1 RNase or pancreatic RNase; they were able to sequence most, but not all, of the oligonucleotides derived from complete T_1 RNase or pancreatic RNase digestion of these fragments. Based on our own results, which include the total sequence of all the oligonucleotides present in complete T_1 RNase and pancreatic RNase digests of the TYMV RNA fragment we have isolated, and based on the overlaps present among large RNA fragments according to the recently revised data of Hirth

and collaborators (104; L. Hirth, personal communication), we may propose a tentative primary sequence for the 3'-115 nucleotides of TYMV RNA (Fig. 39). It should be noted that the two pancreatic RNAse-generated oligonucleotides present in lowest yield in our fingerprints, A8 and A13 (Table V), are not included in this sequence, as they were not present in any of the fragments analyzed by Hirth and collaborators. This is an unfortunate residual ambiguity in the proposed sequence.

One possible secondary structure for the 3'-end of TYMV RNA, derived from the proposed sequence, is that shown in Fig. 39. In formulating this structural model, consideration was given to the guidelines for stability of RNA secondary structures suggested by Tinoco, et al. (171). The only structural element in Fig. 39 not consistent with those rules is stem b (see Fig. 1 for nomenclature). However, such stems may be stabilized by tertiary interactions in the folded molecule and do occur in known tRNA sequences; the similar sequences in stem b of yeast $tRNA_1^{Gly}$ and yeast $tRNA_1^{Asp}$ (11) may be cited as examples.

It is interesting to note that the structure shown in Fig. 39 has many tRNA-like features, although it clearly does not quite fit into a classical cloverleaf pattern (Fig. 1). In the first place, the structure contains the valine anticodon sequence C-A-C (which would read the valine codon GUG) in a plausible 7-membered loop II. This is not inconsistent with the results of Haenni et al. on TYMV valyl-RNA

Figure 39. tRNA-like "cloverleaf" conformation for the 3'-115 nucleotides of TYMV RNA. Arrow indicates most frequent site of cleavage by "RNase P". Possible codon and anticodon sequences are also designated. Further discussion in text.


(A)
. . . U U C C A C C U A A G
Ser
Thr
Ochre
anti-Val

interaction with ribosome-bound RNA copolymers (71), although they favored the codon recognition assignment GUC over GUG. Their results, however, may have been complicated by transacylation of valine, as discussed above. It is also worth noting that the two nucleotides succeeding the anticodon C-A-C in the 5'⟶3' direction are A-C; this is identical with the sequence of both E. coli and yeast $tRNA^{Val}$ (11). This region of loop II has been implicated as crucial to $tRNA^{Val}$ recognition by E. coli valyl-tRNA synthetase (173). It is also interesting that stem e in Fig. 39 resembles the stem e sequence of yeast $tRNA^{Val}$ (11), although it differs from that of E. coli $tRNA^{Val}$ (11). The C-C-A 3'-end of the structure shown in Fig. 39 is preceded by an A, which is consistent with the "Discriminator" hypothesis of tRNA recognition by aminoacyl-tRNA synthetase (174).

The primary sequence we have proposed, and the secondary structure we have derived from it (Fig. 39), must be considered only tentative. Nevertheless, they do point out some potentially interesting features of the 3'-end of TYMV RNA, and suggest experimental approaches that might usefully be applied to further analysis of the RNA's tRNA-like aspects. For example, there are several unusual features in the structure shown in Fig. 39. Especially disturbing is the very small loop IV structure (cf. Fig. 1). The possibility must be considered that loop III of Fig. 39 actually corresponds to the usual tRNA loop IV (Fig. 1), and the sequence

G-U-G-C-A therefore corresponds to the tRNA loop IV sequence G-U-U-C-purine. The first (5'⟶3') U of loop III in Fig. 39 may then possibly interact with the fifth A of this loop, corresponding to the tertiary structural interaction between T_{54} and A_{58} (Fig. 1) reported for yeast $tRNA^{Phe}$ (3,175). Another tertiary structural interaction reported for the T-Ψ-C loop IV sequence of yeast $tRNA^{Phe}$ is the interaction of Ψ-C with the invariant G_{18}-G_{19} of loop I (7,175); if loop III of Fig. 39, lacking a U-U-C sequence, corresponds to the usual loop IV of tRNA, then it is perhaps not surprising that the constant G-G sequence of loop I is also missing. It is interesting that the sequence of the TYMV RNA 3'-fragment loop III (Fig. 39) is idential to the loop IV sequence of several Staphylococcal species' cell wall biosynthesis $tRNA_1^{Gly}$ (11), and is similar to the loop IV sequence reported for wheat germ $tRNA_{1B}^{Gly}$, U-G-C-G-m^1A-U-U (176). In the case of wheat germ $tRNA_{1B}^{Gly}$, the first U of loop IV can be methylated by the ribothymidine methylase of *E. coli* (176). This suggests that the problem of whether it is loop III or loop IV of the structure of Fig. 39 which actually corresponds functionally to loop IV of tRNA might be investigated by using tRNA modifying enzymes as a probe. It has been shown, for example, that TMV RNA can serve as a substrate for *E. coli* ribothymidine methylase and then yields, upon T_1 RNase digestion, a methylated tetranucleotide which co-chromatographs with T-Ψ-C-G (78). An analogous experiment could be

performed with TYMV RNA. It might also be of interest to examine the possible action of m^1A methylase (see position 58, Fig. 1) on TYMV RNA.

Our proposed structure for the 3'-end region of TYMV RNA (Fig. 39) also raises the obvious and intriguing question of how this structure is recognized by and interacts with ValRS. One approach that has proved useful in studying the nature of tRNA interaction with aminoacyl-tRNA synthetase is crosslinking of bound tRNA with the synthetase protein by UV irradiation (177-179). Budzik *et al.* (178) and Shoemaker *et al.* (179) have reported on regions of several tRNAs that can be crosslinked to yeast and to *E. coli* ValRS. Although it is tempting to try to relate their data to ValRS recognition of TYMV RNA, no simple comparisons are possible. The UV crosslinking reaction depends on the proximity of pyrimidine residues in the RNA to cysteine, tyrosine or serine residues in the protein (180), a situation which may not pertain at every site of close interaction; and, in any case, the crosslinking reaction is a measure of close apposition of enzyme and bound substrate, rather than of recognition *per se*. Despite these reservations, however, it might be quite interesting to examine experimentally the regions of the TYMV RNA fragment that might crosslink to ValRS upon exposure to UV irradiation and to compare this with the crosslinking pattern of $tRNA^{Val}$.

C. Function of the tRNA - like End of TYMV RNA

At this time, there is no definitive evidence on the function of the tRNA-like 3'-end of plant viral RNAs. However, the nature of analogous end structures in several phage RNAs and our limited knowledge of plant viral physiology provide ample food for speculation on this point.

Small, plus-stranded RNA viruses of bacteria, plants and animals must use the same RNA genome both as template for replicative RNA polymerases and as messenger for ribosomal protein synthesis. Their RNA genomes may, therefore, be expected to have primary, secondary and tertiary structural "signal" elements for each of these purposes. This is a useful context in which to consider the tRNA-like 3'-structure of plant viral RNAs.

Among small RNA viruses, the physiological processes of the small RNA bacteriophages are so far the best understood. The RNA phages are classified into four or five serological groups (181,182); the most familiar are the group I phages, including MS2, f2 and R17, and the group III phages, including Qβ. Considerable primary nucleotide sequence information is available for the RNAs of each of the aforementioned phages (11); it may be noted in particular that they all share the 3'-terminal tetranucleotide sequence C-C-C-A_{OH}, reminiscent of the 3' C-C-A_{OH} end of tRNA. The replicative cycle of these phages, including the structure and temporal appearance of replicative intermediate double-stranded RNA species,

has been analyzed in some detail (183,184). Replication has also been analyzed in vitro. It was early shown that MS2 RNA replicase (185) and Qβ RNA replicase (186) are highly specific for replication of their homologous template RNAs. In addition, the purification of the RNA replicases of Qβ (187) and f2 (188) has proven to be of great value in studying the mechanism and regulation of phage RNA replication.

Highly purified Qβ RNA replicase is a tetrametic protein consisting of one virus-specific polypeptide and three host polypeptides (189,190). The host polypeptides have been identified as the protein synthesis transfer factors, Tu and Ts (191), and the ribosomal protein S_1 (192). The three host factors have been shown not be be required for the basic RNA chain elongation region (193,277,278); the phage subunit must therefore be the "polymerase", while the host proteins may be required, with natural RNA substrates, for template recognition and/or chain initiation (193,278). The purified Qβ replicase retains a high template specificity; it works well with Qβ RNA minus strand, but requires 1 (194,279) or 2 (195) additional host factors for replication of the plus (virion) strand. An additional function of the Qβ replicase is to regulate translation of the phage genome (196,197) and its binding sites at internal positions on the phage RNA have been examined by various means (198-200).

Because two of the host factors that form part of the Qβ replicase complex are Tu and Ts, proteins that normally interact with and recognize tRNA, an attractive hypothesis

regarding the replicase substrate recognition process is that the phage RNA contains regions of "tRNA-like" structure. This would be consistent with the tentative identification of internal sequences in the phage RNA that are recognized by tRNA nucleotidyl transferase (89) and would ascribe a function to the 3' C-C-A sequence. Furthermore, this hypothesis would be consistent with the finding that crude preparations of both MS2 and Qβ replicases have some activity with TYMV RNA (185,186), which our work has confirmed to have a tRNA-like 3'-end. It should be noted, however, that this hypothesis presupposes the recognition specificity conferred on the Qβ core tetrameric replicase by host factor(s) F_1 (and F_2) (194,195). The highly purified tetrameric replicase, though it contains Tu and Ts, is not competent to recognize the plus strand 3'-end and can function only with Qβ minus strand (195); in fact, despite its capacity for discriminating between Qβ minus strand and other naturally occurring RNAs, Qβ core replicase accepts poly(C) and other C-containing random copolymers (201), as well as a variety of so-called "6S" RNAs (202), and also has the capacity to synthesize RNA in the absence of <u>any</u> added template (203). A tentative model for the minimal template recognition specificity of Qβ core tetrameric replicase has been proposed (204).

Based on the above discussion, it is apparent that a functional analogy for the nature of RNA structural recogni-

tion in the process of initiation of replication may be possible between the plant RNA viruses and the RNA bacteriophages. The plant replicases, too, may consist of a viral coded polymerase subunit that interacts with host factors which confer recognition specificity. However, more specific structural analogies between the replicases or substrate RNAs in question should be examined with caution. As discussed, one could argue that Qβ core replicase, in the presence of host factor(s) F_1 (and F_2), recognizes some general elements of tertiary structure surrounding a C-C-A 3'-end. Although this may resemble recognition of some generalized tRNA features, an equally plausible argument could be made that the 3'-end of Qβ RNA also contains rRNA-like features recognized by the S_1 subunit of Qβ replicase (205). In any case, the 3'-end structures of the plant viral RNAs are far more than generalized "tRNA-like" elements, as each one has been shown definitively to accept only one specific amino acid. Such a high degree of specificity surely implies that the structures have evolved to interact with host proteins that recognize specific tRNAs -- most likely, these are the synthetase enzymes themselves, as they are available in abundance. It is therefore surprising that such reasoning has not been applied in attempts to isolate plant viral RNA replicase activities.

Most eukaryotic viral RNA replicase enzymes that have been studied appear to be bound to cytoplasmic membranes in

association with viral RNA template (206-217). In no case have enzymes of this class been purified to homogeneity, although there is evidence that the active enzyme of polio and other picornaviruses may be represented by a virus-coded polypeptide of 58,000 daltons molecular weight (218), possibly in association with 2 to 4 host proteins (209). This situation may be analogous to that of the bacteriophage RNA replicases. Among plant virus RNA replicases, solubilization and removal of endogenous RNA has been achieved for the enzymes (for nomenclature, see Abbreviations) of TMV (211-213), BMV (214), CMV (215), TYMV (216) and TNV (217); a new polypeptide of 35,000 dalton molecular weigh has been found in extracts containing BMV replicase activity (219). More extensive purification of plant viral replicases has been hampered, in all of these cases, by the lability of the enzymes.

In general, the partially purified plant viral replicases have exhibited only a partial preference for homologous template (211,212, 214-217), although one report did claim an increased specificity of TMV replicase observed upon aging of the enzyme solution during storage (213). In no case, unfortunately, has it been attempted to study the fate of host cognate aminoacyl-tRNA synthetase during the viral life cycle, or to add purified synthetase to an in vitro replication reaction; neither has the presence of synthetase been assayed for in the partially purified replicase preparations.

Though the suggested mode of replication initiation, via template 3'-end recognition by cognate aminoacyl-tRNA synthetase, remains an attractive hypothesis, therefore, there is as yet no corroborative evidence to substantiate it.

Another possible role for the tRNA-like 3'-end in viral development might involve enhancement of the competitive advantage of viral messenger in the utilization of the host protein synthetic apparatus. It is now well-established that in various multicomponent plant RNA viruses, the small encapsidated viral RNA component carries the coat protein cistron (220-225). In general, the respective coat protein messengers can compete out the other homologous viral messenger in cell-free protein synthesis reactions (221-225). It is interesting that in the case of BMV it has been shown that the smallest RNA component, component IV, arises from one of the larger RNA components, component III (226), although the coat cistron is not efficiently translated from component III RNA in vitro (222); component IV RNA is consequently not strictly required for infection (226), as it may be generated again in vivo. A possible function of component IV RNA is to serve as an efficient coat protein messenger late in infection. Related regulatory mechanisms may also occur in animal viruses, as in both Sindbis (227) and Semliki Forest (228) virus-infected cells, smaller RNAs with the same polarity as the encapsidated RNA have been found which, unlike the parent molecules, efficiently direct

the synthesis of the corresponding viral structural proteins in vitro (the initial translation product may be a single precursor polyprotein (227,228)). Recently, a similar phenomenon has also been observed in TMV-infected cells; a 750 nucleotide long coat protein messenger RNA has been isolated that corresponds, essentially, to the 3'-end of the viral RNA (229). Others have studied the cowpea strain of TMV and found a similar short coat protein messenger to be encapsidated into small viral particles (230-232). One of these studies also reported the isolation of an intermediate length TMV particle carrying an intermediate length RNA and alluded to the possibility that each of the three isolated RNAs, long, intermediate and short, acted as a monocistronic messenger for its 5'-proximal cistron (232).

A property associated with both the BMV RNA component IV (222) and the TMV short RNA (231) is the ability to suppress the translation of the other homologous viral messengers in vitro. Although it is possible, in these cases, that the proximity of the tRNA-like structure to the respective ribosome binding sites (i.e., as compared to more distant cistrons) contributes to this competitive advantage, there is no evidence, as yet, to support this. Furthermore, as all the TMV messengers necessarily have the same 3'-tRNA structure (albeit at different distances from the respective coding regions), and the BMV 3'-tRNAs are also all very similar (233), the tRNA structure seems an unlikely candidate for

a factor that mediates differences between the homologous messengers. Rather, the competitive advantage of BMV RNA IV, at least, may be due to its particular ribosome binding site (234), the sequence of which (235) is significantly complementary (6 possible Watson-Crick base-pairs) to the 3'-end sequence of ribosomal 18S RNA (236). Nevertheless, the 3'-tRNA structure may be involved in somehow mediating a competitive advantage of viral RNA over endogenous host mRNA; this could involve either suppression of host mRNA translation, enhancement of viral RNA translation, or both. At this time, one can only suggest that more experimentation in this direction may be warranted.

D. Cleavage of TYMV RNA by E. coli RNase P

Our results indicate several characteristics wherein the 4.5S TYMV RNA fragment we have prepared does not resemble other products that have been reported to result from the *in vitro* action of RNase P. In the first place, the TYMV RNA fragment contains a free 5'-hydroxyl end. The reported phosphodiester bond cleavage mode of RNase P, however, whether acting on tRNA precursors (79,237,238), *E. coli* 4.5S RNA precursor (239), or phage ø80 RNA M_3 (239), is to leave a 5'-phosphate group. As most published observations of this sort have used RNase P no further purified than the enzyme used in our work (i.e., through the DEAE-Sephadex chromatography step of Robertson *et al.* (79)), we also have no cause to suspect that our enzyme preparations were singularly

contaminated with phosphomonoesterase activity so as to quantitatively dephosphorylate an initial RNase P cleavage product. A second anomaly in the structure of the TYMV RNA fragment is the heterogeneity in the 5'-nucleotide of material recovered by gel electrophoresis after *in vitro* phosphorylation with [^{32}P]. This heterogeneity likely reflects an ambiguity in cleavage site specificity of the nuclease(s) responsible for generating this fragment; such ambiguity is highly uncharacteristic of RNase P (239).

Examination of our proposed base-pairing scheme for the 5'- and 3'-terminal sequences of the TYMV RNA fragment (Figs. 38,39) reveals the 5'-end, i.e., the cleavage site, to be in what appears as an extended single-stranded region. Although it could be argued that some or all of this region is normally base-paired with sequences further upstream in the viral RNA, analysis of the sequence of at least another 44 preceding nucleotides (104) does not validate this speculation. Bothwell *et al.* (239) have examined several varieties of RNase P substrates and have proposed two substrate specificities for the enzyme. One involves tRNA recognition and results in cleavage to produce a correct tRNA acceptor stem, while the other involves recognition of the junction between a double-stranded and single-stranded region and results in cleavage at the end of the double-stranded region. Our inferred cleavage site on TYMV RNA (Fig. 39), both in location and in microheterogeneity, fits neither of these models.

Furthermore, our data (Fig. 27) indicate that even the approximate locus of cleavage can vary considerably in different fragment preparations.

It therefore appears likely that the TYMV RNA 3'-fragment is generated by enzyme(s) other than RNase P. The enzyme activity involved may be that endonucleolytic activity (or mixture of activities) that acts on the spacer sequences between the tRNAs of multimeric tRNA precursors, and has been designated RNase P_2 (240,241). This enzyme may partially co-purify with RNase P through DEAE-Sephadex chromatography (241) and can cleave a ...CpA... phosphodiester bond in the single-stranded leader region of the E. coli $tRNA_1^{Tyr}$ precursor (240). However, it has been reported to leave a 5'-phosphate rather than a 5'-hydroxyl group (240), and, as mentioned, we have no reason to suspect significant phosphomonoesterase contamination of our RNase P preparations.

Under various conditions of chemically catalyzed degradation of RNA, it has been shown that the ...CpA... phosphodiester bond is generally the most sensitive to hydrolysis (242). This has also been found to be the bond most sensitive to the action of pancreatic RNase (243). In our own laboratory, we have found the ...CpA... sequence to be by far the most favored cleavage site of an endonucleolytic activity contaminating commercial snake venom phosphodiesterase (J. Heckman, B. Baumstark and M. Silberklang, unpublished observations). More particularly, we have found that almost

any RNA substrate treated sequentially with bacterial alkaline phosphatase and polynucleotide kinase will yield, upon 5'-end group analysis, over 50% A (M. Silberklang and R. Lockard, unpublished observations)[1]. We believe this to be due to the substrate preference of a trace endonucleolytic activity contaminating one or both of these two enzyme preparations. Because our RNase P preparations are only partially purified, it is suggested that the same or a similar contaminating E. coli endonucleolytic activity may be responsible for the observed cleavage preference with TYMV RNA. Such a contaminating activity would be expected to be present in very low, variable amounts; this might explain the very low and variable yield of 4.5S fragment and the general degradation of TYMV RNA reported by our collaborators in Paris (Methods, Table I). This explanation is also consistent with the reported generation of a 4.5S RNA fragment from TYMV RNA by the action of the endonuclease activity contaminating commercial snake venom phosphodiesterase (152). An alternative possibility is that, since all our analyses use polynucleotide kinase, a final artifactual cleavage, due to a contaminating endonuclease, occurs only at the kinase reaction step. However, our accumulated experience, over many years, with the in vitro phosphorylation of intact or base-specific-RNase-digested tRNA and intact globin messenger RNA

---

[1] In order to remove the [^{32}P]-labeled products of phosphorylation at cryptic nicks, we routinely repurify, under denaturing conditions, all polynucleotide substrates phosphorylated in vitro (vide ante, Methods).

leads us to discount this possibility as unlikely; furthermore, our data on the 5'-end of the intact TYMV RNA 3'-fragment is entirely corroborated by the nature and molar ratio of oligonucleotides present in T_1 RNase and pancreatic RNase fingerprints of the fragment.

The above critique pertains only to the possible cleavage of TYMV RNA by the tRNA maturation enzyme(s) of E. coli. It is not possible, at this time, to assess the possible "processing" of TYMV RNA to yield a $tRNA^{Val}$-like RNA fragment in vivo in a eukaryotic host cell. The appearance of a new valine-accepting tRNA species in TYMV-infected Chinese cabbage leaves has, in fact, been reported (244), but this observation has apparently not been pursued further (A. Prochiantz, personal communication).

II. Sequencing Strategies with Non-Radioactive RNA

In the course of our work on the primary nucleotide sequence of the TYMV RNA 3'-fragment, and of Euglena gracilis chloroplast $tRNA^{Phe}$, considerable progress has been made in the elaboration of methods of RNA sequence analysis by in vitro [^{32}P]-end-group-labeling techniques. We originally set out to devise sequencing strategies that were efficient of both sample and time and that were accurate and reproducible in terms of sequence information; all of these criteria have been met. At their current level of sophistication, the methods developed provide a very sensitive and efficient

approach to the sequence analysis of RNA substrates that cannot be radiolabeled to high specific activity *in vivo*.

Many RNA species that are difficult to label with [^{32}P] *in vivo* can also be obtained only in small quantities; this has been true, for example, for eukaryotic organellar tRNAs and for eukaryotic messenger RNA. Small quantities of non-radioactive RNA can be sequenced only by *in vitro* amplification of its detectability, which may be achieved through the use of radionuclides at high specific activity. Radioactive label may either be incorporated into enzymatically copied complementary polynucleotide strands or may be incorporated directly into the original RNA substrate.

Although considerable sophistication has recently been introduced into enzymatic copying procedures to make them suitable for RNA sequencing, these methods remain of limited utility. Both *E. coli* DNA polymerase I and AMV reverse transcriptase have been used successfully in the presence of α-[^{32}P]-labeled nucleoside triphosphates to copy RNA substrates into [^{32}P]-labeled complementary DNA strands that can then be sequenced (245-248). Alternatively, the DNA copy can itself be copied, in turn, by RNA polymerase, to make radioactive RNA of the same polarity as the original substrate (249-251). In all these cases, however, the complementary DNA can only be made by elongation of a primer oligonucleotide, necessitating some prior knowledge of the RNA sequence[1]. Furthermore,

[1]For most eukaryotic messenger RNA, this problem is simplified, as one can use an oligo(dT) primer, complementary to the messenger 3'-poly(A) sequence.

the DNA copy, though faithful in sequence, rarely, if ever, extends to the 5'-extremity of the substrate RNA (252). One enzymatic RNA copying reaction which is capable of de novo initiation and therefore possibly not subject to the primer drawback, is the reaction of Qβ replicase with heterologous RNA substrates in the presence of Mn^{++} (253); however, the precise sites of initiation and termination of this reaction have not yet been determined. At this time, the most general methods for sequence analysis of non-radioactive RNA are those that rely on radioactive post-labeling of the original RNA substrate.

Two approaches are possible to the in vitro post-labeling of non-radioactive nucleic acid; the radioactive label may be introduced either into the base moieties at essentially all positions in the polynucleotide chain, or the label may be introduced into the ribose or deoxyribose moieties at the 5' and/or 3' external positions on the chain. Exmaples of pyrimidine and purine base labeling include the iodination of cytosine in nucleic acids with [^{125}I] (254), the exchange of the C-8 hydrogen of purine residues with [^{3}H]H_2O (255), and the additional exchange of the C-5 hydrogen of cytosine and uracil with [^{3}H]H_2O catalyzed by bisulfite (256). Uniform, base-specific labeling, especially with [^{125}I], has been used for RNA fingerprinting (257); however, it has not yet proven possible to devise a sequencing scheme with such techniques. Several end-group labeling procedures, on the other hand, have been successfully adapted to RNA sequencing.

These include [^{3}H]-labeling of periodate-oxidized 3'-end groups with [^{3}H]-isoniazid (258) or [^{3}H]-borohydride (259, 260), [^{32}P]-labeling of oligonucleotide 3'-end groups with polynucleotide phosphorylase (261), and [^{32}P]-labeling of 5'-end groups with polynucleotide kinase (100,101,262); our methods fall into this last category. Both in the determination of nucleotide composition and in the determination of nucleotide sequence, [^{32}P]-end-group labeling methods compare favorably with methods that use [^{3}H]-end-group labeling.

Because of our laboratory's continuing interest in tRNA, we were at first particularly concerned with developing a sensitive means of analysis for modified nucleoside content. Analytical methods relying on UV detection to identify nucleosides or nucleotides (263-267) generally require from 1 to 3 A_{260} units (50-150 μg) of tRNA to assure detection of minor (1%) components. The periodate-borohydride [^{3}H]-post-labeling method of Randerath (268,269) is nearly 2 orders of magnitude more sensitive, requiring, in its most miniaturized form (269), only 0.02-0.04 A_{260} units (1-2 μg) of tRNA. This procedure does have a major drawback, however, in that certain modified bases are unstable to the chemical labeling reactions, while 2'-O-methylated ribonucleosides are not subject to periodate oxidation and therefore cannot be [^{3}H]-labeled with borotritide. The [^{32}P]-labeling procedure we have developed for modified nucleoside composition

analysis is applicable to 0.0025-0.01 A_{260} units (0.125-0.5 μg) of tRNA, and is thus even more sensitive than the periodate-borotritide method. With the addition of the final nuclease P_1 step, which converts mononucleoside 3'-,5'-diphosphates to the 5'-monophosphate form (see Results), the nucleotide resolution provided by two-dimensional TLC has proved adequate to identify all modified nucleotides examined thus far (Results); in addition, O-methylated nucleotides can be identified unambiguously (see Results). The flexibility of our two-dimensional TLC analysis as to choice of first dimension solvent system (see Results) also allows confirmation of all nucleotide identifications in several solvent systems. The only major drawback of our procedure is its reliance on enzymatic phosphorylation; our results indicate that polynucleotide kinase exhibits considerable substrate preferences. Although further experiments to improve the extent of enzymatic phosphorylation of modified nucleotides are still in progress, it is possible that our analytical procedure will remain only semiquantitative. For critical analyses, therefore, our procedure and the periodate-borotritide procedure may be considered complimentary. Using both procedures, identification of any modified nucleoside in an RNA molecule should be possible, particularly if a UV reference compound is available for corroboration of chromatographic mobility.

As is the case for nucleotide composition analysis, the methods we have described for nucleotide sequence analysis

by [^{32}P]-end-group-labeling appear to be more sensitive than analogous methods that use [^{3}H] post-labeling (259,260). Using the borotritide procedure, Randerath et al. (270),have succeeded in characterizing the oligonucleotides derived from complete T_1 RNase or pancreatic RNase digestions of 3 A_{260} units (60 μg) of tRNA. We have done similar analyses with as little as 0.01 A_{260} units (0.5 μg) of the TYMV RNA fragment. More recently, Randerath and colleagues (271) have presented a preliminary description of procedures incorporating both [^{3}H] and [^{32}P]-post-labeling methodologies to achieve similar analytical sensitivity. However, a persistent drawback of their sequencing methods is the laborious chromatographic procedures on which sequence identifications are based. One the other hand, our use of two-dimensional homochromatography for analysis of partial nuclease digests of [^{32}P]-end-group-labeled oligonucleotides has proven to be a relatively simple and efficient means of sequence analysis. In addition, combining the use of nuclease P_1 and snake venom phosphodiesterase to prepare partial digests of oligonucleotides has provided a general means of sequencing end group labeled oligonucleotides containing modified bases (see Results).

The application of endonuclease digestion with nuclease P_1 to [^{32}P]-end-group-labeled RNA of high molecular weight has also proven a very versatile addition to our repertoire of sequencing strategies. With occasional exceptions (see

Results), we have found that carefully executed two-dimensional homochromatography of such digests yields patterns capable of providing relatively unambiguous sequence information. Sequences of over 20 nucleotides can often be "read" from a single two-dimensional homochromatogram. Our data also show that even RNAs as large as viral genomes can be sequenced directly if they can be cleanly end-labeled. These methods have proven generally useful in many ongoing RNA sequencing projects in our laboratory. In addition to the work described in this thesis, these methods have aided in the completion of the nucleotide sequence analysis of Neurospora crassa $tRNA_i^{Met}$ (20), as well as in the determination of the 5'-proximal nucleotide sequences of rabbit α and β globin messenger RNA (272).

The only comparable strategy for the determination of end-proximal sequences of non-radioactive high molecular weight RNA that has been described is that of Hunt (258), which uses [^{3}H]-isoniazid to label periodate-oxidized 3'-end groups. This method has been used by Shine and Dalgarno to determine the sequence of eight nucleotide stretches at the 3'-ends of prokaryotic 16S ribosomal RNA (273) and eukaryotic 18S ribosomal RNA (236). However, the isoniazid method is considerably less sensitive than that which we have described and does not yield as much sequence information. Furthermore, two incorrect sequences have been published that relied on this method (274,275) raising doubts as to its accuracy.

As methods of sequence analysis by *in vitro* radioactive post-labeling continue to evolve, newer procedures of increased sensitivity and efficiency continue to be introduced. At the inception of this project, we purposed to develop methods of sequence analysis to deal with RNA substrates that are difficult to label with [^{32}P] *in vivo*. The use of nuclease P_1 as a sequencing tool, however, has also suggested other applications of end group labeling to RNA sequence analysis, such as the direct determination of end-proximal sequences of high molecular weight RNA. At this point in our progress with this work, we may already predict that in certain cases, even RNA molecules which *can* be radiolabeled *in vivo* might profitably rather be sequenced by these means.

REFERENCES

1. Crick, F.H.C. (1955) as reproduced in Hoagland, M.B. (1960) in The Nucleic Acids, V. III, ed. Chargaff, E. and Davidson, J.N. (Academic Press, N.Y.), 400-401.
2. 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.
3. Rich, A., and Rajbhandary, U.L. (1976) Ann. Rev. Biochem., 44, 805-860.
4. Quigley, G.J., Wang, A.H., Seeman, N.C., Suddath, F.L., Rich, A., Sussman, J.L., and Kim, S.H. (1975) Proc. Nat. Acad. Sci. USA, 72, 4866-4870.
5. Quigley, G.J., Seeman, N.C., Wang, A.H., Suddath, F.L., and Rich, A. (1975) Nuc. Ac. Res., 2, 2329-2341.
6. Sussman, J.L., and Kim, S.H. (1976) Biochem. Biophys. Res. Commun., 68, 89-96.
7. Ladner, J.E., Jack, A., Robertus, J.D., Brown, R.S., Rhodes, D., Clark, B.F.C., and Klug, A. (1975) Proc. Nat. Acad. Sci. USA, 72, 4414-4418.
8. Ladner, J.E., Jack, A., Robertus, J.D., Brown, R.S., Rhodes, D., Clark, B.F.C., and Klug, A. (1975) Nuc. Ac. Res., 2, 1929-1937.
9. Fraser, T.H., and Rich, A. (1975) Proc. Nat. Acad. Sci. USA, 72, 3044-3048.
10. Sprinzl, M., and Cramer, F. (1975) Proc. Nat. Acad. Sci. USA, 72, 3049-3053.

11. Barrell, B.G., and Clark, B.F.C. (1975) Handbook of Nucleic Acid Sequences (Joynson-Bruvvers Ltd., Oxford, England).

12. Strominger, J. L. (1970) Harvey Lect., 64, 179-213.

13. Riddle,D. L., and Carbon, J. (1973) Nature New Biol., 242, 230-234.

14. Simsek, M. and Rajbhandary, U. L. (1972) Biochem. Biophys. Res. Commun., 49, 508-515.

15. Piper, P. W., and Clark, B. F. C. (1974) Nature, 247, 516-518.

16. Simsek, M., RajBhandary, U. L., Boisnard, M., and Petrissant, G. (1974) Nature, 247, 519-520.

17. Ghosh, K., Ghosh, H. P., Simsek, M., and RajBhandary, U. L. (1974) J. Biol. Chem., 249, 4720-4729.

18. Gillum, A.M., Urquhart, N., Smith, M., and RajBhandary, U. L. (1975) Cell, 6, 395-405.

19. Gillum, A.M., Roe, B. A., Anandaraj, M. P. J. S., and Rajbhandary, U. L. (1975) Cell, 6, 407-413.

20. Gillum, A.M. (1975) Ph. D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

21. Holmquist, R., Jukes, T.S., and Pangburn, S. (1973) J. Mol. Biol., 78, 91-116.

22. Wong, J. T. F. (1975) Proc. Nat. Acad. Sci. USA, 72, 1909-1912.

23. Haseltine, W. A., and Block, R. (1973) Proc. Nat. Acad. Sci. USA, 70, 1564-1568.

24. Brenchley, J. E., and Williams, L. S. (1975) Ann. Rev. Microbiol., 251-274.

25. Soffer, R. L. (1974) Adv. Enzymol., 40, 91-140.

26. Goebel, W., and Helinski, D. R. (1970) Biochemistry, 9, 4793-4801.

27. Lehman, I.R., Roussos, G.G., and Pratt, E.A. (1962) J. Biol. Chem., 237, 829-833.

28. Jacobson, K.B. (1971) Nature New Biol., 231, 17-19.

29. Mischke, D., Kloetzel, P., and Schwochau, M. (1975) Nature, 255, 79-80.

30. McClain, W.H., Guthrie, C., and Barrell, B.G. (1972) Proc. Nat. Acad. Sci. USA, 69, 3703-3707.

31. Nierlich, D.P., Lamfrom, H., Sarabhai, A., and Abelson, J. (1973) Proc. Nat. Acad. Sci. USA, 70, 179-182.

32. Scherberg, N.H., and Weiss, S.B. (1970) Proc. Nat. Acad. Sci. USA, 69, 1114-1118.

33. Wilson, J. (1973) J. Mol. Biol., 74, 753-757.

34. Sueoka, N., and Kano-Sueoka, T. (1970) Prog. Nuc. Ac. Res. Mol. Biol., 10, 23-55.

35. Kano-Sueoka, T., and Sueoka, N. (1969) Proc. Nat. Acad. Sci. USA, 62, 1229-1236.

36. Marchin, G.L., Comer, M.M., and Neidhardt, F.C. (1972) J. Biol. Chem, 247, 5132-5145.

37. Marchin, G.L., Muller, U.R., and Al-Khateeb, G. (1974) J. Biol. Chem., 249, 4705-4711.

38. Comer, M.M., and Neidhardt, F.C. (1975) Virology, 67, 395-403.

39. Scherberg, N.H., and Weiss, S.B. (1970) Proc. Nat. Acad. Sci. USA, 67, 1164-1171.

40. Chen, M., Shiau, R.P., Hwang, L.T., Vaughan, J., and Weiss, S.B. (1975) Proc. Nat. Acad. Sci. USA, 72, 558-562.

41. Henckes, G., Panayotakis, O., and Heyman, T. (1976) J. Virol., 17, 316-325.

42. Subak-Sharpe, H., Shepherd, W.M., and Hay, J. (1966) Cold Spring Harbor Symp. Quant. Biol., 31, 583-594.

43. Morris, V.L., Wagner, E.K., and Roizman, B. (1970) J. Mol. Biol., 52, 247-263.

44. Bell, D., Wilkie, N.M., and Subak-Sharpe, J.H. (1971) J. Gen. Virol., 13, 463-475.

45. Fisher, T.N., and Fisher, E. (1973) Proc. Soc. Exp. Biol. Med., 143, 208-211.

46. Travnicek, M. (1968) Biochym. Biophys. Acta, 166, 757-759.

47. Erikson, R.L. (1969) Virology, 37, 124-131.

48. Baluda, M.A., and Nayak, D.P. (1970) Proc. Nat. Acad. Sci. USA, 66, 329-336.

49. Erikson, E., and Erikson, R.L. (1970) J. Mol. Biol., 52, 387-390.

50. Bishop, J.M., Levinson, W.E., Quintrell, N., Sullivan, D., Fanshier, L., and Jackson, J. (1970) Virology, 42, 182-195.

51. Carnegie, J.N., Deeny, A. O'C., Olson, K.C., and Beaudreau, G.S. (1969) Biochim. Biophys. Acta, 190, 274-284.

52. Waters, L.C., Mullin, B.C., Bailiff, E.G., and Popp, R.A. (1975) J. Virol., 16, 1608-1614.

53. Erikson, E. and Erikson, R.L. (1971) J. Virol., 8, 254-256.

54. Tooze, J. (1974) The Molecular Biology of Tumor Viruses (Cold Spring Harbor Laboratory).

55. Harada, F., Sawyer, R.C. and Dahlberg, J.E. (1975) J. Biol. Chem., 250, 3487-3497.

56. Waters, L.C., Mullin, B.C., Ho, T., Yang, W.K. (1975) Proc. Nat. Acad. Sci. USA, 72, 2155-2159.

57. Waters, L.C. (1975) Biochem. Biophys. Res. Commun., 65, 1130-1136.

58. Dahlberg, J.E., Sawyer, R.C., Taylor, J.M., Faras, A.J., Levinson, W.E., Goodman, H.M., and Bishop, J.M. (1974) J. Virol., 13, 1126-1133.

59. Faras, A.J., Dahlberg, J.E., Sawyer, R.C., Harada, F., Taylor, J.M., Levinson, W.E., Bishop, J.M., and Goodman, H.M. (1974) J. Virol., 13, 1134-1142.

60. Panet, A., Haseltine, W.A., Baltimore, D., Peters, G., Harada, F., and Dahlberg, J.E. (1975) Proc. Nat. Acad. Sci. USA, 72, 2535-2539.

61. Dahlberg, J.E., personal communication to U.L. RajBhandary.

62. Haenni, A.L., Benicourt, C., Teixeira, S., Prochiantz, A., and Chapeville, F. (1975) in Organization and Expression of the Viral Genome - Molecular Interactions in Genetic Translation, Proc. 10th FEBS Meeting, Paris, ed. Chapeville, F. and Grunberg-Manago, M. (North Holland, American Elsevier) 39, 121-131.

63. Beljanski, M. (1965) Bull. Soc. Chim. Biol., 47, 1645-1646.

64. Pinck, M., Yot, P., Chapeville, F. and Duranton, H.M. (1970) Nature, 226, 954-956.

65. Yot, P., Pinck, M., Haenni, A.L., Duranton, H.M., and Chapeville, F. (1970) Proc. Nat. Acad. Sci. USA, 67, 1345-1352.

66. Kossel, H., and RajBhandary, U.L. (1968) J. Mol. Biol., 35, 539-560.

67. Litvak, S., Carre, D.S., and Chapeville, F. (1970) FEBS Lett., 11, 316-319.

68. Prochiantz, A., and Haenni, A.L. (1973) Nature New Biol., 241, 168-170.

69. Litvak, S., Tarrago, A., Tarrago-Litvak, L., and Allende, J.E. (1973) Nature New Biol., 241, 88-90.

70. Haenni, A.L., Prochiantz, A. and Yot, P. (1974) in Lipmann Symposium: Energy, Biosynthesis and Regulation in Molecular Biology, ed. D. Richter (Walter deGruyter, Berlin, New York), 264-276.

71. Haenni, A.L., Prochiantz, A., Bernard, O., and Chapeville, F. (1973) Nature New Biol., 241, 166-168.

72. Pinck, M., Chan, S.K., Genevaux, M., Hirth, L., and Duranton, H. (1972) Biochimie, 54, 1093-1094.

73. Pinck, L., Genevaux, M., Bouley, J.P., and Pinck, M. (1975) Virology, 63, 589-590.

74. Hall, T.C., Shih, D.S., and Kaesberg, P. (1972) Biochem. J., 129, 969-976.

75. Kohl, R.J., and Hall, T.C. (1974) J. Gen. Virol., 25, 257-261.

76. Oberg, B., and Philipson, L. (1972) Biochem. Biophys. Res. Commun., 48, 927-932.

77. Benicourt, C., and Haenni, A.L. (1974) FEBS Lett., 45, 228-231.

78. Marcu, K., and Dudock, B. (1975) Biochem. Biophys. Res. Commun., 62, 798-807.

79. Robertson, H.D., Altman, S., and Smith, J.D. (1972) J. Biol. Chem. 247, 5243-5251.

80. Pinck, M., Genevaux, M., and Duranton, H. (1974) Biochimie, 56, 423-428.

81. Chen, J.M., and Hall, T.C. (1973) Biochemistry, 12, 4570-4574.

82. Shih, D.S., Kaesberg, P., and Hall, T.C. (1974) Nature, 249, 353-355.

83. Sela, I. (1972) Virology, 49, 90-94.

84. Salomon, R., Sela, I., Soreq, H., Giveon, D., and Littauer, U.Z. (1976) Virology, 71, 74-84.

85. Salomon, R., and Littauer, U.Z. (1974) Nature, 249, 32-34.

86. Stebbing, N., personal communication to A.L. Haenni.

87. Spector, D.H. and Baltimore, D. (1975) J. Virol., 16, 1081-1084.

88. Gillespie, D., Takemoto, K., Robert, M., and Gallo, R.C. (1973) Science, 179, 1328-1330.

89. Prochiantz, A., Benicourt, C., Carre, D. and Haenni, A.L. (1975) Eur. J. Biochem., 52, 17-23.

90. Yogo, Y., and Wimmer, E. (1972) Proc. Nat. Acad. Sci. USA, 68, 1877-1882.

91. Pinkerton, T.C., Paddock, G., and Abelson, J. (1972) Nature New Biol., 240, 88-90.

92. Paddock, G., and Abelson, J. (1973) Nature New Biol., 246, 2-6.

93. Ohe, K., and Weissman, S.M. (1970) Science, 167, 879-881.

94. Ohe, K., Weissman, S.M., and Cooke, N.R. (1969) J. Biol. Chem., 244, 5320-5332.

95. Guilley, H., Jonard, G., and Hirth, L. (1975) Proc. Nat. Acad. Sci. USA, 72, 864-868.

96. Lamy, D., Jonard, G., Guilley, H., and Hirth, L. (1975) FEBS Lett., 60, 202-204.

97. Briand, J.P., Richards, K.E., Bouley, J.P., Witz, J., and Hirth, L. (1976) Proc. Nat. Acad. Sci. USA, 73, 737-741.

98. Kaesberg, P., in Animal Virology, vol. IV, ICN/UCLA Symposia, ed. Baltimore, D., and Huang, A.S., (Academ. Press) 555-565.

99. Brownlee, G.G. (1972) Determination of Sequences in RNA (North-Holland, Amsterdam).

100. Szekely, M., and Sanger, F. (1969) J. Mol. Biol., 43, 607-617.

101. Simsek, M., Ziegenmyer, J., Heckman, J., and RajBhandary, U., L. (1973) Proc. Nat. Acad. Sci. USA, 70, 1041-1045.

102. Walker, R.T. and RajBhandary, U.L. (1975) Nuc. Ac. Res., 2, 61-78.

103. Sanger, F., Donelson, J.E., Carlson, A.R., Kossel, H., and Fischer, D. (1973) Proc. Nat. Acad. Sci. USA, 70, 1209-1213.

104. Briand, J.P., Jonard, G., Guilley, H., Richards, K., and Hirth, L. (1976) Eur. J. Biochem., submitted.

105. Fujimoto, M., Kuninaka, A., and Yoshino, H. (1974) Agr. Biol. Chem., 38, 1555-1561.

106. Panet, A., van de Sande, J.H., Loewen, P,C., Khorana, H.G., Raae, A.J., Lillehaug, J.R., and Kleppe, K. (1973) Biochemistry, 12, 5045-5050.

107. Bishop, D.H.L., Obijeski, J.F., and Simpson, R.W. (1971) J. Virol., 8, 66-73.

108. Hall, R.H., and Khorana, H.G. (1955) J. Amer. Chem. Soc., 77, 1871-1875.

109. Moffatt, J.G., and Khorana, H.G. (1961) J. Amer. Chem. Soc., 83, 663-675.

110. Zishka, M.K., and Nishimura, J.S. (1970) Anal. Biochem., 34, 291-297.

111. Soll, D., Cherayil, J., Jones, D.S., Faulkner, R.D., Hampel, A., Bock, R.M., and Khorana, H.G. (1966) Cold Spring Harbor Symp. Quant. Biol., 31, 51-61.

112. Peacock, A.C., and Dingman, C.E. (1967) Biochemistry, 6, 1818-1827.

113. Akroyd, P. (1967) Anal. Biochem., 19, 399-410.

114. Studier, F.W. (1973) J. Mol. Biol., 79, 237-248.

115. Knecht, M.E., and Busch, H. (1971) Life Sci., 10, 1297-1309.

116. DeWachter, R., and Fiers, W. (1972) Anal. Biochem., 49, 184-197.

117. Turchinsky, M.F., and Shereshneva, L.P. (1973) Anal. Biochem. 54, 315-318.

118. Southern, E.M., and Mitchell, A.R. (1971) Biochem, J., 123, 613-617.

119. Pearson, R.L., Weiss, J.F., and Kelmers, A.D. (1971) Biochim. Biophys. Acta, 228, 770-774.

120. Leberman, R. (1966) Virology, 30, 341-347.

121. Gillam, I., Blew, D., Warrington, R.C., von Tigerstrom, M., and Tener, G.M. (1968) Biochemistry, 7, 3459-3468.

122. Demushkin, V.D., Belitova, O.D., and Budovskii, E.I. (1971) Mol. Biol. (U.S.S.R.), 5, 689-692.

123. Schofield, P., Hoffman, B.M., and Rich, A. (1970) Biochemistry, 9, 2525-2533.

124. Glynn, I.M., and Chapell, J.B. (1964) Biochem. J., 90, 147-149.

125. Symons, R.H. (1968) Biochim. Biophys. Acta, 155, 609-610.

126. Padmanabhan, R., and Wu, R. (1972) J. Mol. Biol., 65, 447-467.

127. Wimmer, E. (1972) J. Mol. Biol, 68, 537-540.

128. Barrell, B.G. (1971) in Procedures in Nucleic Acid Research v. 2, ed. G.L. Cantoni and D.R. Davies (Harper and Row, New York) 751-779.

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

130. Novogrodsky, A., Tal, M., Traub, A., and Hurwitz, J. (1965) J. Biol. Chem., 241, 2933-2943.

131. Richardson, C.C. (1965) Proc. Nat. Acad. Sci. USA, 54, 158-165.

132. Van de Sande, J.H., and Bilsker, M. (1973) Biochemistry, 12, 5056-5062.

133. Dube, S.K., Marcker, K.A., Clark, B.F.C., and Cory, S. (1969) Eur. J. Biochem., 8, 244-255.

134. Walker, R.T., and RajBhandary, U.L. (1972) J. Biol. Chem., 247, 4879-4892.

135. Nishimura, S., Harada, F., Narushima, U., and Seno, T. (1967) Biochim. Biophys. Acta, 142, 133-148.

136. Becker, A., and Hurwitz, J. (1967) J. Biol. Chem., 242, 936-950.

137. Hall, R.H. (1971) The Modified Nucleosides in Nucleic Acids (Columbia University Press, New York).

138. Lillehaug, J. R., and Kleppe, K. (1975) Biochemistry 14, 1225-1229.

139. Khorana, H. G. (1968) Biochem. J., 109, 709-725.

140. Wu, R., (1970) J. Mol. Biol., 51, 501-521.

141. Razzel, W. E. (1963) Meth. Enzymol., 6, 236-258.

142. Ling, V. (1972) J. Mol. Biol., 64, 87-102.

143. Ziff, E. B., Sedat, J. W., and Galibert, F. (1973) Nature New Biol., 241, 34-37.

144. Rensing, V. F. E., and Schoenmakers, J. G. G. (1973) Eur. J. Biochem., 33, 8-18.

145. Bambara, R., Jay, E., and Wu, R. (1974), Nuc. Ac. Res., 1, 1503-1520.

146. Tu, C. P. D., Jay, E., Bahl, C. P., and Wu, R. (1976) Anal. Biochem., 74, 73-93.

147. Handbook of Biochemistry (1970), ed. H. A. Sober (Chemical Rubber Co., Cleveland, Ohio), Sect. G, p. 88.

148. Murray, K., and Murray, N. (1973) Nature New Biol., 243, 134-139.

149. Loewen, P. C. (1975) Nuc. Ac. Res., 2, 839-852.

150. Kelly, T. J., and Smith, H. O. (1970) J. Mol. Biol., 51, 393-409.

151. Roberts, R., Catalogue of Restriction Endonucleases, unpublished.

152. Litvak, S., Tarrago-Litvak, L. and Chapeville, F. (1973) J. Virol., 11, 238-242.

153. Chang, S. H., Hecker, L. I., Silberklang, M., Brum, C. K., Barnett, W. E., and RajBhandary, U. L. (1976) Fed. Proc. Abstracts, 35, 150.

154. Chang, S.H., Brum, C.K., Silberklang, M., RajBhandary, U.L., Hecker, L.I., and Barnett, W.E. (1976) Cell, in press.

155. Hecker, L.I., Uziel, M., and Barnett, W.E. (1976) Nuc. Ac. Res., 3, 371-380.

156. Wintermeyer, W., and Zachau, H.G. (1970) FEBS Lett., 11, 160-164.

157. RajBhandary, U.L., and Chang, S.H. (1968) J. Biol. Chem., 243, 598-608.

158. Ohashi, Z., Maeda, M., McCloskey, J.A., and Nishimura, S. (1974) Biochemistry, 13, 2620-2625.

159. Friedman, S., Li, H.J., Nakanishi, K., and van Lear, G. (1974) Biochemistry, 13, 2932-2936.

160. Erdmann, V.A., Fahnestock, S., Higo, K., and Nomura, M. (1971) Proc. Nat. Acad. Sci. USA, 68, 2932-2936.

161. Erdmann, V.A., Sprinzl, M., and Pongs, O. (1973) Biochem. Biophys. Res. Commun., 54, 942-948.

162. Richter, D., Erdmann, V.A., and Sprinzl, M. (1974) Proc. Nat. Acad. Sci. USA, 71, 3226-3229.

163. von der Haar, F., Schlimme, E., Erdmann, V.A., and Cramer, F. (1971) Bioorganic Chem., 1, 282-293.

164. Boedtker, H., and Kelling, D.G. (1967) Biochem. Biophys. Res. Commun., 29, 758-766.

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

166. Rushizky, G.W., and Sober, H.A. (1962) J. Biol. Chem., 237, 834-840.

167. Smith, J.D. (1976) Prog. Nuc. Ac. Res. Mol. Biol., 16, 25-73.

168. Burdon, R.H. (1974) Brookhaven Symposium: Processing of RNA, 26, 138-153.

169. Horowitz, J., Ou, C.N., Ishaq, M., Ofengand, J. and Bierbaum, J. (1974) J. Mol. Biol., 88, 301-312.

170. Ofengand, J., Bierbaum, J., Horowitz, J., Ou, C.N., and Ishaq, M., (1974) J. Mol. Biol., 88, 313-325.

171. Tinoco, I., Uhlenbeck, O.C., and Levine, M.D. (1971) Nature, 230, 362-367.

172. Stehelin, D., Peter, R., and Duranton, H. (1973) Biochim. Biophys. Acta, 293, 253-265.

173. Mirzabekov, A.D., Lastity, D., Levina, E.S., and Bayev, A.A. (1971) Nature New Biol., 229, 21-22.

174. Crothers, D.M., Seno, T. and Soll, D.G. (1972) Proc. Nat. Acad. Sci. USA, 69, 3063-3067.

175. Kim, S.H., Suddath, F.L., Quigley, G.J., Wang, A.H.J., Seeman, N.C., and Rich, A. (1974) Science, 185, 435-440.

176. Marcu, K., Mignery, R., Reszelbach, R., Roe, B., Sirover, M., and Dudock, B. (1973) Biochem. Biophys. Res. Commun., 55, 477-483.

177. Shoemaker, H.J.P., and Schimmel, P.R. (1974) J. Mol. Biol., 84, 503-513.

178. Budzik, G.P., Lam, S.S.M., Shoemaker, H.J.P., and Schimmel, P. (1975) J. Biol. Chem., 250, 4433-4439.

179. Shoemaker, H.J.P., Budzik, G.P., Giege, R., and Schimmel, P.R. (1975) J. Biol. Chem., 250, 4440-4444.

180. Smith, K.C., and Meun, H.C. (1968) Biochemistry, 7, 1033-1037.

181. Sakurai, T. T., Miyake, T., Shiba, T., and Watanabe, I. (1968) Jpn. J. Microbiol., 12, 544-546.

182. Miyake, T. I., Haruma, I., Shiba, T., Itoh, Y. H., Yamane, K., and Watanabe, I. (1971) Proc. Nat. Acad. Sci. USA, 68, 2022-2024.

183. Zinder, N. D. (1968) Harvey Lect., 62, 1-25.

184. Stavis, R. L., and August, J. T. (1970) Ann. Rev. Biochem., 39, 527-560.

185. Haruna, I., Nozu, K., Ohtaka, Y., and Spiegelman, S. (1963) Proc. Nat. Acad. Sci. USA, 50, 905-911.

186. Haruna, I., and Spiegelman, S. (1965) Proc. Nat. Acad. Sci. USA, 54, 579-587.

187. Kamen, R. (1972) Biochim. Biophys. Acta, 262, 88-100.

188. Federoff, N. V., and Zinder, N. D. (1971) Proc. Nat. Acad. Sci. USA, 68, 1838-1843.

189. Kamen, R. (1970) Nature, 228, 527-533.

190. Kondo, M., Gallerani, R., and Weissmann, C. (1970) Nature, 525-527.

191. Blumenthal, T., Landers, T. A., and Weber, K. (1972) Proc. Nat. Acad. Sci. USA, 69, 1313-1317.

192. Wahba, A. L., Miller, M. J., Niveleau, A., Landers, T. A., Carmichael, G. G., Weber, K., Hawley, D. A., and Slobin, L. I. (1974) J. Biol. Chem., 249, 3314-3316.

193. Landers, T. A., Blumenthal, T., and Weber, K. (1974) J. Biol. Chem., 249, 5801-5808.

194. Kamen, R. I., Monstein, H. J., and Weissmann, C. (1974) Biochim. Biophys. Acta, 366, 292-299.

195. Franze deFernandez, M. T., Haywood, W.S., and August, J. T. (1972) J. Biol. Chem., 247, 824-831.

196. Kolakofsky, D., and Weissmann, C. (1971) Nature New Biol. 231, 42-46.

197. Kolakofsky, D., and Weissmann, C. (1971) Biochim. Biophys. Acta, 246, 596-599.

198. Weber, H., Kamen, R., Meyer, F., and Weissmann, C. (1974) Experientia, 30, 711.

199. Meyer, F., Weber, H., Vollenweider, H. J., and Weissmann, C. (1975) Experientia, 31, 743.

200. Vollenweider, H. J., Koller, T., Weber, H., and Weissman, C. (1976) J. Mol. Biol., 101, 367-377.

201. Franze de Fernandez, M. T., Eoyang, L., and August, J. T. (1968) Nature, 219, 588-590.

202. Banerjee, A.K., Rensing, U., and August, J. T. (1969) J. Mol. Biol., 45, 181-193.

203. Kacian, D. L., Mills, D. R., Kramer, F. R. and Spiegelman, S. (1972) Proc. Nat. Acad. Sci. USA, 69, 3038-3042.

204. Kuppers, B., and Sumper, M. (1976) Proc. Nat. Acad. Sci. USA, 72, 2640-2643.

205. Engelberg, H., and Schoulaker, R. (1976) J. Mol. Biol., in press.

206. Girard, M., Baltimore, D., and Darnell, J. E. (1967) J. Virol., 10, 504-515.

207. Caliguiri, L. A., and Mosser, A. G. (1971) Virology, 46, 375-386.

208. Rosenberg, H., Diskin, B., Oron, L., and Traub, A. (1972) Proc. Nat. Acad. Sci. USA, 69, 3815-3819.

209. Friedman, R.M., Levin, J.G., Grimely, P.M., and Berezesky, I.K. (1972) J. Virol., 10, 504-515.

210. Lazarus, L.H., and Barzilae, R. (1974) J. Gen. Virol., 23, 213-218.

211. Zaitlin, M., Duda, C.T., and Petti, M.A. (1973) Virology, 53, 300-311.

212. Brishammar, S., and Junti, N. (1974) Virology, 59, 245-253.

213. Sela, I., and Hauschner, A. (1975) Virology, 64, 284-288.

214. Hadidi, A., and Fraenkel-Conrat, H. (1973) Virology, 52, 363-372.

215. Clark, G.L., Peden, K.W.C., and Symons, R.H. (1974) Virology, 62, 434-443.

216. Mouches, C., Bove, C., and Bove, J.M. (1974) Virology, 58, 409-423.

217. Fraenkel-Conrat, H. (1976) Virology, 72, 23-32.

218. Lundquist, R.E., Ehrenfeld, E., and Maizel, J.V. (1974) Proc. Nat. Acad. Sci. USA, 71, 4773-4777.

219. Hariharasubramanian, V., Hadidi, A., Singer, B., and Fraenkel-Conrat, H. (1973) Virology, 54, 190-198.

220. Bol, J.F., van Vloten-Doting, L., and Jaspars, E.M.J. (1971) Virology, 46, 73-85.

221. Mohier, E., Hirth, L., LeMeur, M.A., and Gerlinger, P. (1975) Virology, 68, 349-359.

222. Shih, D.S., and Kaesberg, P. (1973) Proc. Nat. Acad. Sci. USA, 70, 1799-1803.

223. Davies, J.W., and Kaesberg, P. (1974) J. Gen. Virol., 25, 11-20.

224. Schwinghamer, M.W., and Symons, R.H. (1975) Virology, 63, 252-262.

225. Mayo, M.A., Fritsch, C., and Hirth, L. (1976) Virology, 69, 408-415.

226. Shih, D.S., Lane, L.C., and Kaesberg, P. (1972) J. Mol. Biol., 64, 353-362.

227. Simmons, D.T., and Strauss, J.H. (1974) J. Mol. Biol., 86, 397-409.

228. Clegg, C., and Kennedy, I. (1975) Eur. J. Biochem., 53, 175-183.

229. Hunter, T.R., Hunt, T., Knowland, J., and Zimmern, D. (1976) Nature, 260, 759-764.

230. Whitfeld, P.R., and Higgins, T.J.V. (1976) Virology, 71, 471-485.

231. Higgins, T.J.V., Goodwin, P.B., and Whitfeld, P.R. (1976) Virology, 71, 486-497.

232. Bruening, G., Beachy, R.N., Scalla, R., and Zaitlin, M. (1976) Virology, 71, 498-517.

233. Bastin, M., Dasgupta, R., Hall, T.C., and Kaesberg, P. (1976) J. Mol. Biol., 103, 737-745.

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

235. DasGupta, R., Shih, D.S., Saris, C., and Kaesberg, P. (1975) Nature, 256, 624-628.

236. Shine, J., and Dalgarno, L. (1974) Biochem. J., 141, 609-615.

237. Carbon, J., Chang, S., and Kirk, L.L. (1974) Brookhaven Symposium in Biology: Processing of RNA, 26, 26-36.

238. Barrell, B.G., Seidman, J.G., Guthrie, C., and McClain, W.H. (1974) Proc. Nat. Acad. Sci. USA, 71, 413-416.

239. Bothwell, A.L.M., Stark, B.C., and Altman, S. (1976), Proc. Nat. Acad. Sci. USA, 73, 1912-1916.

240. Schedl, P., Primakoff, P., and Roberts, J. (1974) Brookhaven Symposium: Processing of RNA, 26, 53-76.

241. Schedl, P., Roberts, J., and Primakoff, P. (1976) Cell, 8, 581-594.

242. Singer, B., and Fraenkel-Conrat, H. (1963) Biochim. Biophys. Acta, 76, 143-145.

243. Witzel, H., and Barnard, E. (1962) Biochem. Biophys. Res. Commun., 7, 295-298.

244. Yot, P., Czosnek, H., and Haenni, A.L. (1971) Abstr. Commun. Seventh Meet. Eur. Biochem. Soc., 141.

245. Milstein, C., Brownlee, G.G., Cartwright, E.M., Jarvis, J.M., and Proudfoot, N.J. (1974) Nature, 252, 354-359.

246. Proudfoot, N.J., and Brownlee, G.G. (1974) Nature, 252, 359-362.

247. Proudfoot, N.J. (1976) Nuc. Ac. Res., 3, 1811-1821.

248. Salser, W., Bowen, S., Browne, D., El Adli, F., Federoff, N., Fry, K., Heindell, H., Paddock, G., Poon, R., Wallace, B., Whitcome, P. (1976) Fed. Proc., 35, 23-35.

249. Marotta, C.A., Forget, B.G., Weissman, S.M., Verma, I.M., McCaffrey, R.P., and Baltimore, D. (1974) Proc. Nat. Acad.

Sci. USA, 71, 2300-2304.

250. Poon, R., Paddock, G. V., Heindell, H., Whitcome, P., Salser, W., Kacian, D., Bank, A., Gambino, R., and Ramirez, F. (1974) Proc. Nat. Acad. Sci. USA, 71, 3502-3506.

251. Paddock, G. V., Heindell, H. C., and Salser, W. (1974) Proc. Nat. Acad. Sci. USA, 71, 5017-5021.

252. Maniatis, T., Kea, S. G., Efstratiadis, A., and Kafatos, F. C. (1976) Cell, 8, 163-182.

253. Palmenberg, A., and Kaesberg, P. (1974) Proc. Nat. Acad. Sci. USA, 71, 1371-1375.

254. Commerford, S. L. (1971) Biochemistry, 10, 1993-2000.

255. Tomasz, M., Olson, J., and Mercado, C. M. (1972) Biochemistry, 11, 1235-1241.

256. I Ida, S., Wataya, Y., Kudo, I., Kai, K., and Hayatsu, H. (1974) FEBS Lett., 39, 263-266.

257. Robertson, H. D., Dickson, E., Model, P., and Prensky, W. (1973) Proc. Nat. Acad. Sci. USA, 70, 3260-3264.

258. Hunt, J. A. (1970) Biochem. J., 120, 353-363.

259. Randerath, K., Randerath, E., Chia, L. S. Y., Gupta, R. C., and Sivarajan, M. (1974) Nuc. Ac. Res., 1, 1121-1141.

260. Sivarajan, M., Gupta, R. C., Chia, L. S. Y., Randerath, E., and Randerath, K. (1974) Nuc. Ac. Res., 1, 1329-1341.

261. Szeto, K. S., and Soll, D. (1974) Nuc. Ac. Res., 1, 171-181.

262. Slack, J. M., and Loening, U. E. (1974) Eur. J. Biochem., 43, 59-67.

263. Uziel, M., Koh, C., and Cohn, W. E. (1968) Anal. Biochem. 25, 77-98.

264. Singhal, R. P. (1972) Arch. Biochem. Biophys., 152, 800-810.

265. Sen, G. C., and Ghosh, H. P. (1974) Anal. Biochem., 58, 578-591.

266. Rogg, H., Brambilla, R., Keith, G., and Staehelin, M. (1976) Nuc. Ac. Res., 3, 285-295.

267. Nishimura, S. (1972), Prog. Nuc. Ac. Res. Mol. Biol., 12, 49-85.

268. Randerath, E., Yu, C. T., and Randerath, K. (1972) Anal. Biochem., 48, 172-198.

269. Randerath, K., and Randerath, E. (1973), Meth. Cancer Res., 9, 3-69.

270. Randerath, K., Chia, L. S. Y., Gupta, R. C., Randerath, E., Hawkins, E. R., Brum, C. K., and Chang, S. H. (1975) Biochem. Biophys. Res. Commun., 63, 157-163.

271. Gupta, R. C., Randerath, E., and Randerath, K. (1976) Fed. Proc. Abstracts, 35, 149.

272. Lockard, R. E., and RajBhandary, U. L. (1976) Cell, in press.

273. Shine, G., and Dalgarno, L. (1975) Nature, 254, 34-38.

274. Hunt, J. A. (1973) Biochem. J., 131, 315-325.

275. Hunt, J. A., and Oakes, G. N. (1976) Biochem. J., 155, 637-644.

276. Altman, S., and Smith, J. O. (1971) Nature New Biol., 233, 35-39.

277. Kamen, R., Kondo, M., Romer, W., and Weissmann, C. (1972) Eur. J. Biochem., 31, 44-51.

278. Carmicheel, G. G., Landers, T. A., and Weber, K. (1976) J. Biol. Chem., 251, 2744-2748.

279. Carmichael, G. G., Weber, K., Niveleau, A., and Wahba, A. J. (1975) J. Biol. Chem., 250, 3607-3612.