Volume 13 Number 12 1985 Nucleic Acids Research

Structure and expression of the M_2 genomic segment of a type 2 killer virus of yeast

Ernest M.Hannig and Michael J.Leibowitz

Department of Microbiology, University of Medicine and Dentistry of New Jersey, Rutgers Medical School, P.O. Box 101, Piscataway, NJ 08854, USA

Received 21 February 1985; Revised and Accepted 23 May 1985

ABSTRACT

The M_2 double-stranded (ds) RNA species encodes toxin and resistance functions in Saccharomyces cerevisiae strains with the K_2 killer specificity. RNA sequence analysis reveals the presence of a large open reading frame on the larger heat-cleavage product of M_2 dsRNA, which is translated in vitro to yield a 28 kd polypeptide as a major product. The postulated translation initiator AUG triplet is located within a stem and loop structure near the 5'terminus of the positive strand, which also contains plausible 18S and 5.8S ribosomal RNA binding sites. These features may serve to regulate the translation of the K_2 toxin precursor. The M_1 (from type 1 yeast killers) and M_2 dsRNA species lack extensive sequence homology, although specific features are shared, which may represent structural elements required for gene expression and replication.

INTRODUCTION

The killer systems of Saccharomyces cerevisiae provide an opportunity to study both virus/host and virus/virus interactions in a genetically well-defined eukaryotic system. S. cerevisiae killer strains secrete a protein toxin which is lethal to sensitive strains, but against which the killers bear specific resistance. The genetic information for toxin production (K^+ phenotype) and resistance (R^+ phenotype) is present on virion-encapsidated, linear double-stranded (ds) RNA molecules denoted M. A number of different killer specificities exist, with the K_1 (containing M_1 dsRNA) and the K_2 (containing M_2 dsRNA) specificities being the most widely studied. The maintenance of M_1 dsRNA (1830 base pairs) is dependent upon at least 27 chromosomal MAK (maintenance of killer) genes, in addition to PET18, SPE2, and SPE10. Mutations in certain other genes, denoted SKI (super killer), allow the maintenance of M_1 dsRNA in certain Mak^- strains (reviewed in 1,2,3).

All killer strains, as well as many non-killers, contain a larger 4.9 kilobase pair (kbp) viral dsRNA species denoted L-A (4), which appears to encode the major capsid polypeptide of M- and L-A-containing particles (4-9). Several non-Mendelian genes are located on L-A dsRNA. These include:

i) [HOK], which allows M_1 dsRNA to replicate in a Ski^+ host (10);

ii) [EXL], whose presence results in the exclusion of M_2 dsRNA (11);
iii) [NEX], which prevents the exclusion of M_2 dsRNA by [EXL] (11).
In addition, the presence of [NEX] results in the exclusion of M_2 dsRNA in an *mkt* host (maintenance of [KIL-k_2]) at 30°C, but not at 20°C (11,12). Other virion-encapsidated dsRNAs of the same size as L-A, denoted L-B and L-C, have no known direct interaction with the killer system (4,9,13). L-A- and M_1-containing virus particles co-purify with a transcriptase activity which catalyzes the synthesis of full-length positive polarity transcripts of L-A and M_1 dsRNAs, denoted 1 and m_1 respectively (7,14,15). L-B- and L-C-containing particles can be separated from M_1 and L-A particles, and have an RNA polymerase activity which is biochemically distinct from that of the particles harboring M_1 and L-A (9). The maintenance of L-A dsRNA requires some, but not all, of the genes required for M_1 dsRNA maintenance (2,16).

The M_1 dsRNA from K_1 killer yeasts can be cleaved, either by S_1 nuclease or heat treatment, into two fragments denoted M_1-1 (1000 bp) and M_1-2 (630 bp) (17). Cleavage occurs at an internal A,U-rich region (18) of variable length (Hannig and Leibowitz, in preparation), which consists of primarily adenine residues on the positive-polarity strand (19). *In vitro*, denatured M_1-1 (17) and M_1 (20) dsRNAs, like the m_1 transcript (17), program the synthesis of a 32-34 kilodalton (kd) preprotoxin, which is processed *in vivo* to produce mature toxin and (presumably) resistance factor (21-23). Translation of the preprotoxin initiates at the first AUG (bases 14-16) from the 5' terminus of the positive strand (19,22,24,25), and may be regulated by the presence of secondary structure (19,25) and potential 18S and 5.8S ribosomal RNA binding sites (19,24-26) present near the 5' terminus of the positive strand. M_1-1, and thus the entire preprotoxin gene, has been sequenced from a cDNA copy (22,27). M_1 dsRNA is transcribed *in vivo* to produce both full-length and subgenomic (1.2 kb) positive polarity transcripts which bind to poly(U)-Sepharose as tightly as do polyadenylated RNAs with a 3'-terminal poly(A) tract of 70-100 residues (19,28). Identical behavior has been demonstrated for the full-length positive polarity m_1 transcript, synthesized *in vitro*, which is not 3'-terminally polyadenylated, but rather contains the internal A-rich region characteristic of the genomic positive strand (19). This has led to the hypothesis of template-coding of the poly(A)-like tract of the *in vivo* transcripts (19,29).

In this report, we characterize the M_2 dsRNA from K_2 yeast killers. Maintenance of M_2 dsRNA (1550 bp) is dependent upon the *MAK8*, *MAK10*, *MAK16*, *MKT1*, *MKT2*, *SPE2*, and *SPE10* chromosomal genes (2,16). Presumably, others are also required. Crossing K_1 and K_2 killers yields only K_1 diploids which have retained M_1 dsRNA, but

have lost M_2 dsRNA (30). The exclusion of M_2 dsRNA by the presence of M_1 dsRNA represents a third mode of M_2 dsRNA exclusion, in addition to that mediated by [EXL] or by [NEX] in an mkt host (31). We have characterized M_2 dsRNA with respect to overall structure, terminal RNA sequences, in vitro transcription and translation, and secondary structure of the coding strand. Although strikingly similar to M_1 dsRNA in overall structure, the primary sequence of M_2 dsRNA differs drastically from that of M_1 dsRNA with the exception of a few blocks of conserved sequences. In addition, we have identified a coding region for a potential precursor for the K_2 toxin.

MATERIALS AND METHODS

S. cerevisiae strains, cell growth, dsRNA extraction and virion purification.

S. cerevisiae strain 1384 (αhis4 [KIL-k_2] [HOK] [NEX-o] [EXL-o])(11) used in these experiments was kindly provided by R.B. Wickner. Double-stranded RNA (15,18) or transcriptionally active virions (9) were prepared from cells grown to late stationary phase (5 days, 28°C, 150 rpm) in YPE medium consisting of 1% (w/v) yeast extract, 2% (w/v) peptone and 3% (v/v) ethanol.

In vitro transcription of strain 1384 virions.

Transcription reactions were performed in vitro, as described previously (15), in the presence or absence of [α-^{32}P]-UTP (Amersham Corp.). In some cases (see text), m_2 and 1 transcripts were purified, following phenol extraction and ethanol precipitation, by oligo(dT)-cellulose (Collaborative Research, Type 3) chromatography (19,32) or electrophoresis through a 1% agarose gel (19).

Analysis of in vitro translation products

RNA species were translated in vitro in a rabbit reticulocyte lysate protein synthesis system (Amersham Corp., batch 26) in the presence of [^{35}S]- methionine (Amersham Corp.). Marker [^{14}C]-proteins were from New England Nuclear. Double-stranded RNAs, purified as above, were boiled for three minutes in sterile glass-distilled water prior to translation. The m_2 transcript, synthesized in vitro, was purified by oligo(dT)-cellulose chromatography prior to translation. Final reaction mixtures (20 ul) contained 73 mM K^+ and 1.22 mM Mg^{2+} cations, and were incubated at 30°C for 75 minutes. Reactions were terminated by freezing at -20°C. Prior to gel electrophoresis (33), samples were treated with RNase A as described previously (17). Radioactive species were visualized by fluorography of the gels (34) using Kodak XAR-5 film.

Sequence analysis of M_2 double-stranded RNA

The 3' termini of M_2 dsRNA were reacted with [5'-^{32}P]-pCp (Amersham Corp.) in a reaction catalyzed by bacteriophage T_4 RNA ligase (Pharmacia/P-L Biochemicals,

RNase-free), as described previously (24). RNA species bearing a single radioactive terminus were purified by two cycles of polyacrylamide gel electrophoresis following heat cleavage (17, 24) or strand separation (25) of the radioactive dsRNA. For strand separation, M_2 dsRNA was subjected to electrophoresis at 7-8 v/cm for 16-18 hours. This is a lower voltage gradient than that which yields optimal strand separation of M_1 dsRNA (25). Both 3' termini incorporated [5'-^{32}P]-pCp with approximately equal efficiency. The 5' termini of M_2 dsRNA were labeled using [γ-^{32}P]-ATP (Amersham Corp.) as described previously (19), followed by two cycles of strand separation gel electrophoresis (see above). The efficiency of incorporation of ^{32}P at the 5' terminus of the positive strand was approximately 10-fold greater than incorporation at the 5' terminus of the negative strand, as has been observed for M_1 dsRNA (19).

Chemical (35) and enzymatic (19,25) sequence analyses were performed as described previously. The specificities of the enzymes used for sequence analysis are as follows: ribonuclease T_1 (G), pancreatic ribonuclease (C+U), ribonuclease PhyM (A+U), ribonuclease CL-3 (C), and ribonuclease U_2 (A) as cited (19,25). Digestion products were analyzed on 20%, 8%, or 5% polyacrylamide gels containing 8 M urea (36). Formamide ladders were generated as described previously (37). Terminal base analyses (5' and 3') were performed as described (19,38).

Secondary structure analysis of 5'-[^{32}P] positive strands

Positive strands of M_2 dsRNA bearing 5'-[^{32}P] were purified by two cycles of strand separation electophoresis as described above. The positive strands were digested with either ribonuclease T_1 (Calbiochem-Behring Corp.) at 1.67 or 16.7 U/ml, or S_1 nuclease (Sigma) at 16.7 or 167 U/ml (39,40). Digestions were performed in the presence of 0.17 M NaCl and 10 ug *E. coli* B tRNA (GIBCO). Reaction mixtures (final volume 6 ul) were incubated at 37°C for ten minutes in 1.5 ml polypropylene tubes, chilled on ice for 1-2 minutes and, following the addition of 1 ul of the appropriate enzyme, incubated at 37°C for an additional ten minutes. Controls were treated similarly, but without the addition of enzyme. Reactions were terminated by the addition of 2 ul of 50 mM Na_2EDTA, followed by freezing in a dry ice/ethanol bath. An equal volume of a sample buffer (10 M urea, 0.05% bromphenol blue and xylene cyanol) was added, and digestion products were analyzed on urea-polyacrylamide gels as described above.

Denaturing gel electrophoresis and blot hybridization analysis

RNA samples were denatured and subjected to electrophoresis through 1.2% agarose gels containing formaldehyde (41). For blot hybridization analysis, RNA was transferred to nitrocellulose paper (42), baked at 78°C for 2-3 hours in a vacuum oven, and hybridized with [^{32}P]-l and/or [^{32}P]-m_2 transcripts synthesized in vitro

in the presence of [α-^{32}P]-UTP (see above). Hybridization conditions (65°C, 5 x SSC, 50% formamide) and washing conditions (final wash at 65°C in 0.25 x SSC-0.1% SDS) have been described (19). Blots were then air dried and exposed to X-ray film at -70°C with or without an intensifying screen (Dupont).

RESULTS

M_2 double-stranded RNA contains an internal A,U-rich region

M_2 dsRNA, like M_1 dsRNA, can be separated into electrophoretic subspecies on native polyacrylamide gels (data not shown). In the case of M_1 dsRNA, these species have identical 3'-terminal sequences (24), and are believed to represent conformers of M_1 dsRNA. Alternatively, this phenomenon could be due, in part, to the variation in the length of the A,U-rich region of M_1 dsRNA (43), although similar subspecies are seen in M_1 dsRNA preparations apparently lacking variability at that site (Hannig and Leibowitz, manuscript in preparation). These observations prompted us to test for the presence of an internal A,U-rich region in M_2 dsRNA.

Treatment of M_2 dsRNA with the single-strand specific S_1 nuclease (Figure 1A), or heat (Figure 1B), produced two fragments which we denote M_2-1 (1.05 kbp) and M_2-2 (ca. 0.37 kbp). The lengths are based upon the mobility of these fragments in native agarose and polyacrylamide gels. These fragments are derived from opposite termini, since they both retain 3'-terminal ^{32}P (Figure 1B), and each contains a unique terminal sequence (see below). Oligo(dT)-cellulose chromatography of separated strands of 3'-[^{32}P]-M_2 dsRNA demonstrates that the faster-migrating strand, which we show to be of positive polarity (see below), binds to the column (>93%). The slower-migrating strand is not bound by the column (>99% unbound), nor is native M_2 dsRNA (>97% unbound). Consistent with the internal location of this A-rich region on the positive strand, oligo(dT)$_{12-18}$ will prime the synthesis of a ca. 1 kb cDNA from denatured M_2 dsRNA by reverse transcriptase (E.M. Hannig, unpublished data). This priming resulted in first strand synthesis with 3-8% efficiency, depending upon conditions; a similar range of efficiency was observed for reverse transcription of denatured M_1 dsRNA.

Polarity of the separated strands of M_2 dsRNA

We compared the polarity of each separated strand of M_2 dsRNA to that of the m_2 transcript, synthesized in vitro by the virion-associated transcriptase, by blot hybridization analysis. The results, shown in Figure 2, demonstrate that the fast strand (Fig.2, lane 2) is of the same polarity as the m_2 transcript, while the slow strand (Fig. 2, lane 1) is complementary to the m_2 transcript. The m_2 transcript also appears to be of full-length size. The probe used for this blot (Fig.2, lane 5) is a phenol-extracted in vitro transcription reaction which was incubated in the

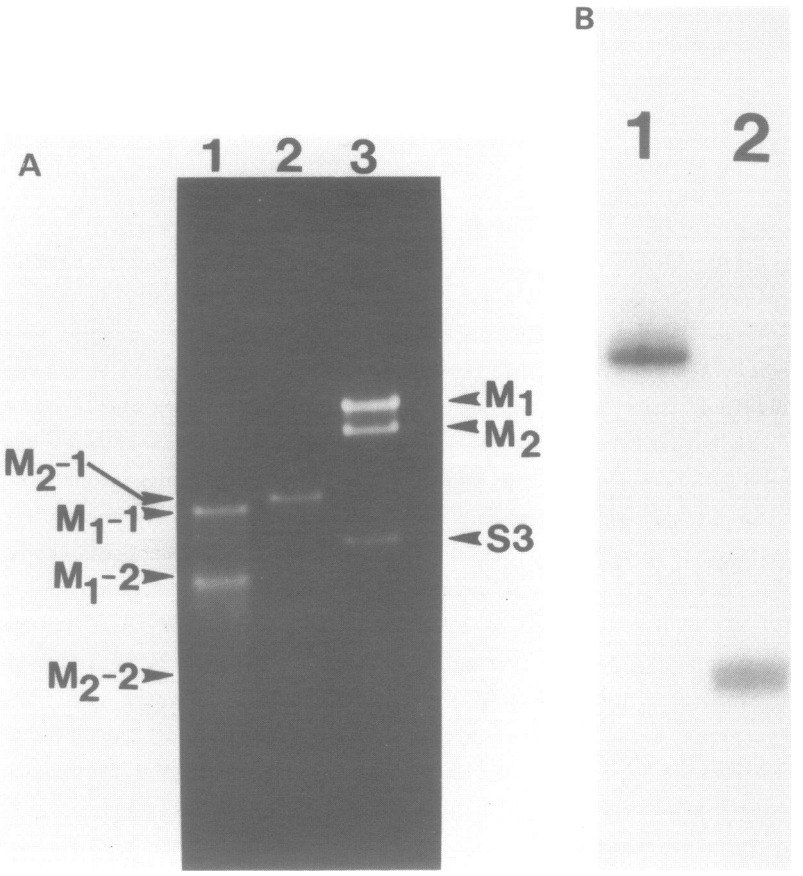


Figure 1. Internal cleavage of M_2 dsRNA. A. Double-stranded RNA (0.4ug) was treated with S1 nuclease and subjected to electrophoresis through a 1.5% agarose gel in the presence of ethidium bromide (1 ug/ml), as described (17). Lane 1 contains M_1 dsRNA from strain A364A x S7 (14), lane 2 contains M_2 dsRNA, and lane 3 contains untreated M_1, M_2 and S3 (730bp; ref. 18) dsRNAs. B. M_2 dsRNA bearing 3'-terminal [5'-^{32}P]-pCp was subjected to heat cleavage and analyzed by electrophoresis through a 5% polyacrylamide gel (24). Purified heat-cleavage products were subjected to a second cycle of polyacrylamide gel electrophoresis and individual RNA species were visualized by autoradiography: lane 1, M_2-1 dsRNA; lane 2, M_2-2 dsRNA.

presence of [α-^{32}P]-UTP and thus contains a mixture of 1 and m_2 transcripts. Identical results were obtained using purified m_2 transcript as the probe (data not shown). Additionally, this experiment demonstrates that, under the relatively stringent hybridization and washing procedures used here, there is no cross-hybridization between denatured M_1 dsRNA and the m_2 (or 1) transcript.

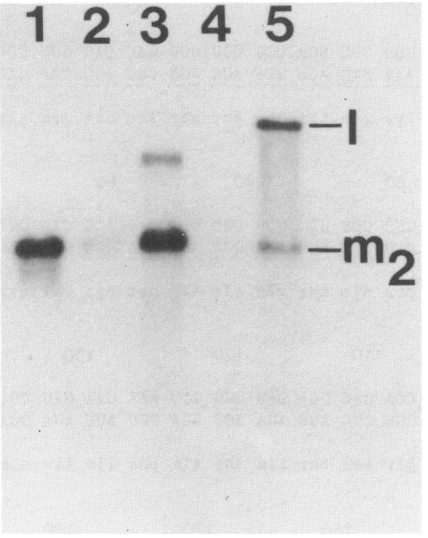


Figure 2. Polarity of M_2 dsRNA separated strands. Purified M_2 dsRNA separated strands and the M_1 and M_2 dsRNAs were denatured and subjected to electrophoresis through a 1.2% agarose gel containing formaldehyde, transferred to nitrocellulose paper and hybridized with [^{32}P]-UMP in vitro transcription products from virions purified from strain 1384. Lane 1, M_2 dsRNA slow strand; lane 2, M_2 dsRNA fast strand; lane 3, M_2 dsRNA; lane 4, M_1 dsRNA; lane 5, probe used for hybridization. The positions of the l and m_2 transcripts, present as part of the probe, are indicated. The low level of hybridization seen in lane 2 is presumably due to the presence of contaminating slow strand. "Fast" and "slow" refer to the relative mobilities of the strands upon strand separation electrophoresis. The larger hybridizing species in lane 3 may be incompletely denatured dsRNA (19,29).

As further evidence of a poly(A)-like tract on positive strands derived from M_2 dsRNA, we analyzed in vitro transcription products, synthesized by virions purified from strain 1384, by oligo(dT)-cellulose chromatography. This procedure separates the l (unbound) from the m_2 (bound) transcripts, although a small amount of agarose gel-purified l transcript (<2%) binds to the column, as is the case for l transcripts derived from virions purified from a K_1 killer (19). Using agarose gel-purified m_2 transcript, we find that only full-length m_2 transcript is bound by oligo(dT)-cellulose (data not shown). However, a heterogeneous group of TCA-precipitable radioactive species present in gel-purified m_2 transcript is not bound by the column, and may represent breakdown of the transcript.

Terminal sequence of M_2-1

We determined the sequence of the termini of M_2-1, the larger heat-cleavage product of M_2 dsRNA, by chemical (3'-terminal ^{32}P label) and enzymatic (3'- or 5'-

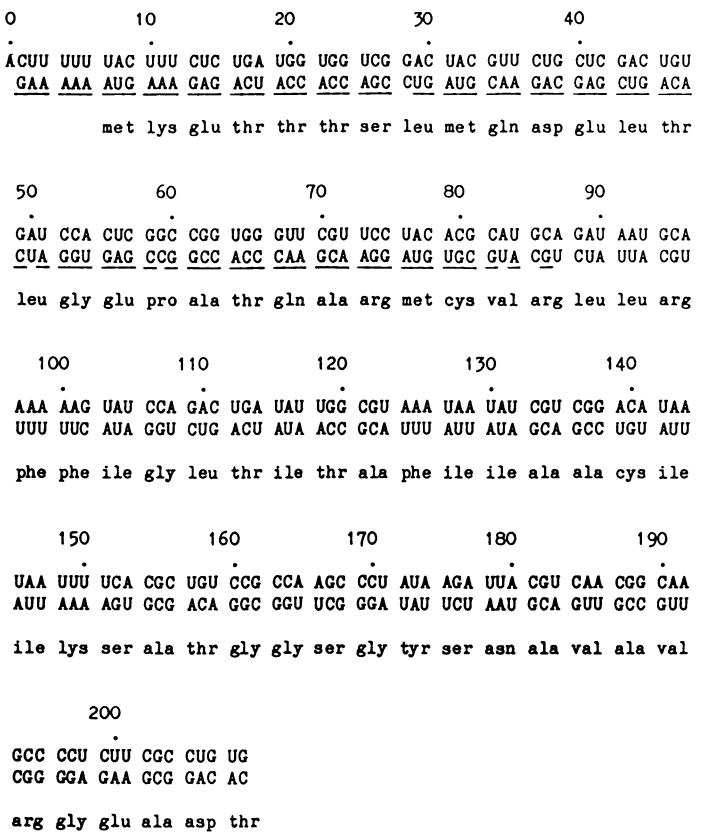


Figure 3. Terminal RNA Sequence of M_2-1. The sequence of the 3'-terminus of the negative strand, as determined from 3'-[^{32}P]-RNA, is shown on the top line. The 3'-terminal unpaired A residue is designated base zero. The bottom line of the nucleotide sequence represents the 5' terminus of the positive polarity strand. Underlined bases on the second line were determined directly from the 5' terminus of the positive strand; those not underlined were predicted from the negative strand sequence. An AUG at positions 7-9 of the positive strand begins an open reading frame predicted to encode a polypeptide with the indicated partial amino acid sequence.

terminal ^{32}P label) procedures. RNA sequences were determined from both heat-cleaved fragments and separated strands. The M_2-1 fragment contains the 5'-terminus of the positive strand, and the 3' terminus of the negative strand. End-group analysis reveals that the 3'-terminus is A(83%)>G(11.6%). The 5'-terminal nucleotide is G(71%)>U(24.5%). Upon sequence analysis, this 5'-terminal heterogeneity was not evident, although faint shadow bands consistent with the presence of low levels of U 5'-terminal to G1 were seen in some digestions (see Figure 6, lane

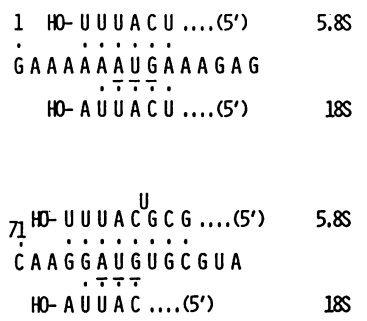


Figure 4. Potential ribosomal RNA binding sites. Possible base-pairing interactions between the 3' termini of yeast 5.8S and 18S rRNAs, and the regions containing the AUG codons at positions 7-9 and 76-78 from the 5' terminus of the M_2 dsRNA positive strand, are indicated. Other pairings are possible, but these are predicted to have greatest stability.

T_1). In contrast, the 5'-terminal nucleotide of the genomic M_1 dsRNA positive strand is G(>97%) (19). As is the case with M_1 dsRNA (24), agarose gel-purified M_2 dsRNA is contaminated with short oligonucleotides that incorporate ^{32}P from [5'-^{32}P]-pCp in an RNA ligase-catalyzed reaction, but which are subsequently removed by polyacrylamide gel electrophoresis.

The sequence data are summarized in Figure 3. The 3'-terminal A residue on the negative strand is unpaired, and is designated base zero. An open reading frame, on the positive strand, beginning at an AUG at nucleotides 7-9 extends for as far as we have sequenced. The +4 and -3 nucleotides are A residues (assigning the A of the AUG as +1), placing this AUG in a favorable initiating context (44). The leader region is rich in A residues and poor in G residues, which is characteristic of many yeast messenger RNAs (45), including that derived from M_1 dsRNA (19,24,25). This 5'-leader, however, is relatively short when compared with other eukaryotic mRNAs (44) and the m_1 transcript (19).

Potential 18S (46) and 5.8S (47) ribosomal RNA binding sites can be drawn near the AUG at bases 7-9. Similar pairings have been postulated to function in the translation of the M_1 preprotoxin (26). The ΔG^o values for the 18S and 5.8S rRNA interactions are -7.4 kcal and -8.6 kcal, respectively (48). The only other AUG in our sequence which fits both the context and potential rRNA binding criteria is located at nucleotides 76-78, and is also in the same reading frame as the first AUG. In this case, the context is relatively poor (a U in the +4 position), and the potential rRNA interactions are relatively weak. The ΔG^o for the postulated 18S and 5.8S rRNA interactions are -4.0 kcal and -4.2 kcal, respectively. These interactions are summarized in Figure 4.

During our sequence analysis, we observed that strong cleavage at bases U3, U10, U31, U46, C52, U56, C78, C80 and U90, numbered from the 3' terminus of the negative strand, occurred in the G-specific chemical sequencing reaction when the 3'-[^{32}P] negative strands were used as a substrate. Chemical sequencing of the M_2-

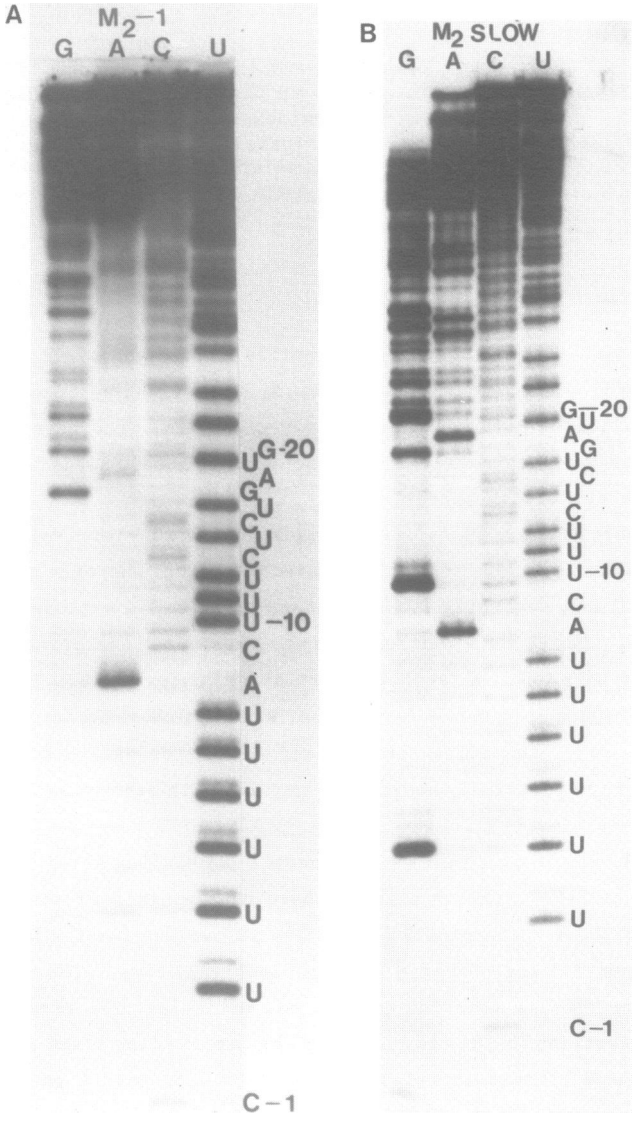


Figure 5. Chemical sequence analysis at the 3' terminus of the M_2 dsRNA negative strand. The M_2-1 fragment (A) or the M_2 dsRNA slow strand (B) was purified from 3'-[^{32}P]-M_2 dsRNA, analyzed by chemical sequence analysis, and resulting products were subjected to electrophoresis through urea-polyacrylamide gels. The C-specific reaction is weak on these 20% polyacrylamide gels. Note the G-specific cleavages at U3 and U10 in (B) which are absent in (A).

1 fragment, derived by heat cleavage of 3'-[^{32}P]-M_2 dsRNA, from the 3' terminus showed cleavage only at the correct pyrimidine nucleotide. This is clearly shown for U3 and U10 in Figure 5. Base C106, numbered as above, displayed this phenomenon in a 30 second chemical G reaction, but not in a 20 second reaction. Therefore, this phenomenon may depend upon length of treatment, but it is difficult to judge based upon a single case. Modification of residues in a single-stranded (separated strands) versus double-stranded (heat cleaved fragments) configuration is an alternative explanation, although the treatment with dimethylsulfate at 90°C, as occurs in the G-specific reaction, should denature the dsRNA as it does for RNA secondary structure (35). These aberrantly reactive pyrmidines are located in single-stranded, or locally unstable, regions of the secondary structure predicted (see below) to exist at the 3'-terminus of the minus strand. However, we observe the occurrence of this phenomenon only at selected pyrimidine residues (perhaps due to sequence context). This was not observed when sequencing from the 3'-terminus of the opposite strand. None of these unusual bases are recognized by G-specific T_1 ribonuclease digestion of 3'-[^{32}P]-negative strands (data not shown). We know of no modified bases that behave in this manner, and are unaware of a published account describing a similar phenomenon.

We confirmed the base-pairing assignments for the unusual pyrimidines, up to U90, by partial enzymatic cleavage of 5'-[^{32}P]-positive strands. A representative sequencing gel is shown in Figure 6. It is clear that, for those bases displayed, all agree with the sequence predicted from the opposite strand. The aberrant pyrimidines base-pair with the predicted purine base, rather than with cytosine which might be expected from the G-specific reactivity of the base on the opposite strand (Fig. 6, and data not shown).

Comparison with M_1-1 and the M_1 preprotoxin

Visual and computer-assisted comparisons of the terminal nucleotide sequences of M_1-1 and M_2-1 reveal little sequence homology. Homologies of six or more nucleotides include (1)GAAAAAU(M_1) and (1)GAAAAA^{A}U(M_2), (22)GCC^{A}ACCCAAG(M_1) and (61)GCCACCCAAG(M_2), and (54)UAUUAUUUUUCAU(M_1) and (89)UAUUA^{CGU}UUUUUCAU(M_2). The numbers in parentheses indicate the beginning nucleotide position, numbered from the 5'-terminus of the positive strand, and the superscript bases are additional sequences not present in one or the other dsRNA species. The first and third homologies seem to be particularly significant, as biochemical evidence indicates that these regions are paired at the base of a stem-and-loop structure present at the 5' terminus of the positive strands of both M_1 (19,25) and M_2 (see below) dsRNAs. The 3'-terminal U rich region complementary to the 5' terminus of the positive strand has been suggested to function in transcriptase recognition (49). The only

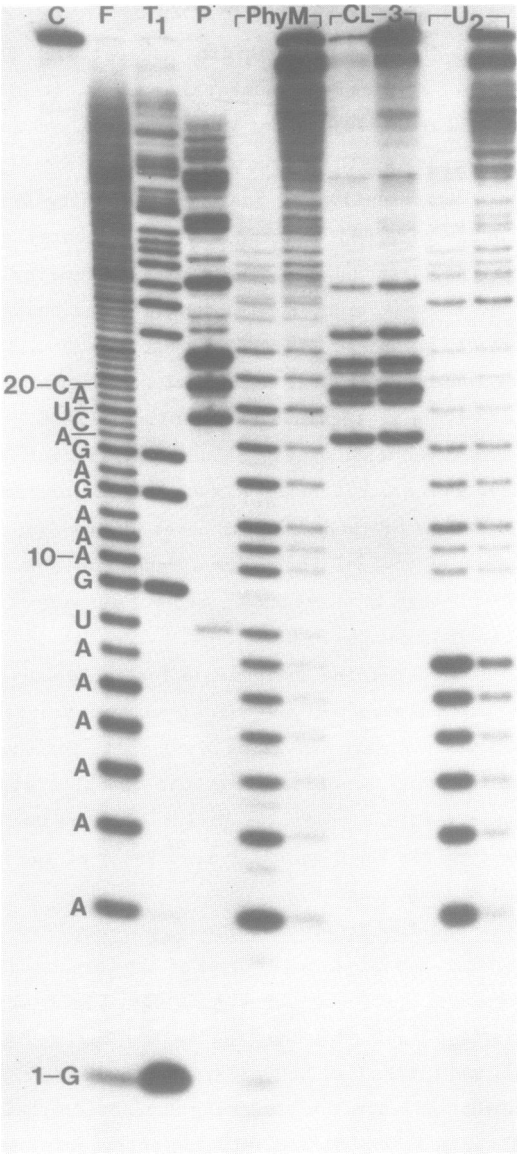


Figure 6. Enzymatic sequence analysis of 5'-[^{32}P]-M_2 dsRNA positive strands. 5'-[^{32}P]-M_2 dsRNA positive strands were purified by two cycles of strand separation electrophoresis, followed by partial digestion with base-specific endoribonucleases. Cleavage products were fractionated, in this case, on a 20% acrylamide-8M urea gel and visualized by autoradiography. Lane C, no enzyme control; lane F, partial formamide hydrolysis; lane P, pancreatic ribonuclease. Other lanes are marked with the corresponding endoribonuclease. Phy M, CL-3 and U_2 nucleases were used at two concentrations, which represent a 6-fold difference in each case.

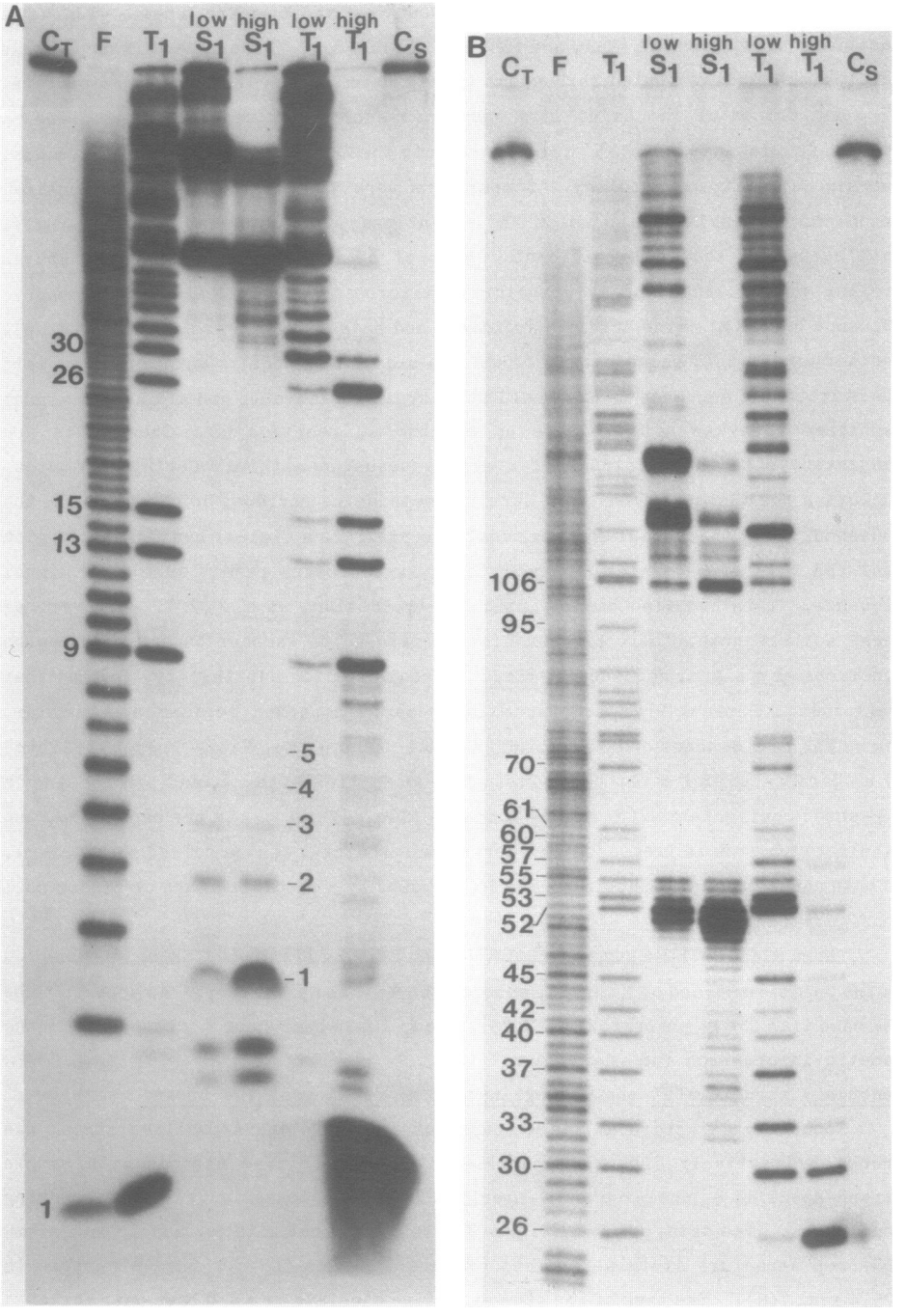
other two seven-nucleotide homologies between M_1-1 and M_2-1 are those beginning at positions 54 (M_1) and 141 (M_2), and 188 (M_1) and 88 (M_2). Nine other six nucleotide homologies are distributed throughout these two molecules in no apparent pattern.

Comparison of the amino acid sequences encoded by the large open reading frames, initiating near the 5' terminus of both the M_1 and M_2 dsRNA positive strands, shows no significant homology. Comparisons were made on the basis of identical amino acids, similar amino acids and evolutionary relatedness. Interestingly, both sequences show a significant region of hydrophobicity (50) which shares similar predicted secondary structural features (51). In M_2, the hydrophobic sequence begins at amino acid 27 (arginine) and ends near amino acid 45 (cysteine). Whether or not this sequence functions as a signal sequence is unknown. However, the hydrophobic sequence is preceeded by an arginine residue, and a possible signal peptidase cleavage site follows an alanine at position 50. These data are consistent with the existence of a signal sequence although certain features, including the distance from the N-terminal methionine residue, do not quite fit the consensus model (52). Our computer analysis predicts a similar hydrophobic region near the N-terminus of the M_1 preprotoxin, which bears resemblance to a signal sequence. This has also been pointed out by Bostian, et al. (22). However, as these authors note, there are conflicting data as to whether or not this region functions as a signal sequence in vivo. Perhaps not too suprisingly, we find that the predicted secondary structures in these hydrophobic regions are similar, beginning in a β-sheet configuration (M_1, Met_1 -Phe_{17}; M_2, Met_{24}-Gly_{34}), followed by a short α-helical stretch (M_1,Ile_{18}-Val_{24}; M_2,Ile_{37}-Cys_{45}), and terminating in a β-sheet configuration (M_1,Val_{25}-Asn_{28}; M_2,Ile_{46}-Thr_{51}). In the case of M_2, the α-helical stretch is preceeded by a very short predicted random coil (Leu_{35},Thr_{36}). In both cases, the hydrophobic region is followed by a predicted turn or random coil configuration.

These similarities suggest that the open reading frame, beginning at bases 7-9 on the positive strand of M_2 dsRNA, encodes the M_2 toxin precursor. As appears to be the case with the M_1 preprotoxin (22,23), the M_2 toxin precursor may also contain the genetic information for resistance.

Secondary structure at the 5'-terminus of M_2 dsRNA positive strand

Secondary structure at the 5'-terminus of the M_2 dsRNA positive strand was probed by partially digesting purified 5'-[^{32}P]-positive strands with single strand-specific S_1 nuclease and G-specific T_1 ribonucleases under non-denaturing conditions. The data, shown in Figure 7, are consistent with our model for a stem-and-loop structure at this terminus (Figure 8). The ΔG^o for this structure is -25.4 kcal (48). The base of the stem structure consists of a G·C base pair followed


A
low high low high
C_T F T_1 S_1 S_1 T_1 T_1 C_S
30
26
15
13
9
1
-5
-4
-3
-2
-1
B
low high low high
C_T F T_1 S_1 S_1 T_1 T_1 C_S
106
95
70
61
60
57
55
53
52
45
42
40
37
33
30
26

by 6 A·U base pairs. This feature is conserved at the 5'-terminus of the M_1 dsRNA positive strand, where there are 8 A·U base pairs (19,25). Cleavage by S_1 nuclease at the first few bases of the stem presumably occurs because the presence of the A·U base pairs allows the molecule to "breathe" in solution. The AUG at nucleotides 7-9 is located in a relatively unstable region of the predicted structure. It begins two bases after the last weakly S_1 nuclease-sensitive nucleotide in the series (Figure 7) and is followed by unpaired A residues, two A·U base pairs and unpaired purine bases. This region also contains potential 18S and 5.8S rRNA binding sites (see above). Thus, in spite of the high degree of base-pairing in this region, the AUG at bases 7-9 could be accessible to the cellular translation machinery. Perhaps such a structure is sufficient, however, to protect the 5'-terminus from degradation by cellular exonucleases. The AUG at positions 76-78 is also located in a locally unstable region of this stem-and-loop structure. Ribosomal RNA base pairing with either of these regions might destabilize the secondary structure predicted for the M_2 positive strand by 3.4 kcal (18S rRNA) or 4.6 kcal (5.8S rRNA) at AUG 7-9, or by 7.4 kcal (18S rRNA) or 11kcal (5.8S rRNA) at AUG 76-78 (48).

In vitro translation of M_2 dsRNA

Both the m_2 transcript (synthesized *in vitro*) and denatured M_2 dsRNA program the synthesis of a major 28 kd [^{35}S]-methionine-containing species, which we denote M-p28, in a rabbit reticulocyte lysate *in vitro* protein synthesis system (Figure 9). Methionine incorporation in at least two larger species (38 kd, 40.5 kd) and two smaller, more heterogenous species (17 kd, 15.5 kd) is directed by these RNAs. Interestingly, a labeled species migrating slightly ahead of globin is visible in the M_1-, M_2- and m_2- programmed reactions. In the latter case, this band is clearly visible on longer exposures (data not shown). Denatured M_2-1 encodes only M-p28 in the reticulocyte lysate system, although we do occasionally also see a species migrating ahead of globin (data not shown). These data suggest, by analogy with M_1 dsRNA (17), that M-p28 represents the M_2-specific killer toxin precursor and is

Figure 7. *Secondary structure analysis of 5'-[^{32}P]-M_2 dsRNA positive strands.* Positive (fast) strands of 5'-[^{32}P]-M_2 dsRNA were purified by two cycles of strand separation electorphoresis, digested with T_1 ribonuclease or S_1 nuclease, and products were subjected to electrophoresis through a 20% (A) or 8% (B) polyacrylamide gel containing 8M urea. Lanes C_T and C_S no enzyme controls for T_1 ribonuclease and S_1 nuclease, respectively; lane F, partial formamide hydrolysis; lane T_1, sequencing T_1 ladder, displaying G residues. The numbers on the left refer to the distance of the indicated G residues from the 5' terminus of the positive strand. Low and high T_1 ribonuclease (G-specific) and S_1 nuclease (non-specific) levels were used to digest the positive strands, in the presence of 0.17 M NaCl, as described in Materials and Methods. Numbers 1-5 in (A) refer to S_1 nuclease cleavage sites. As pointed out previously (25), oligonucleotides from S_1 digestions, migrate with fragments 1 to 2 bases longer in the formamide lane (F).

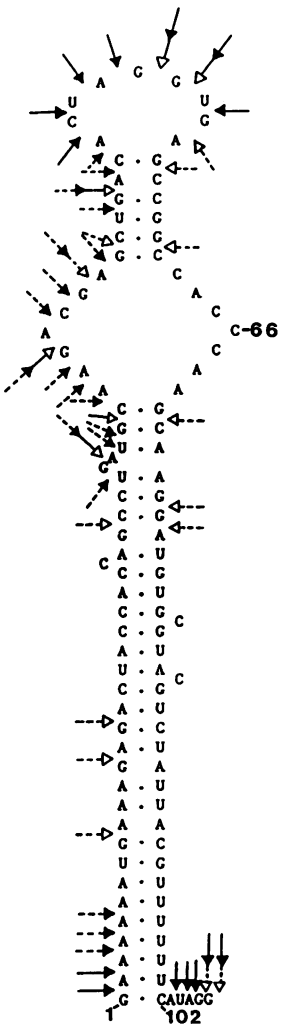


Figure 8. Model for the secondary structure at the 5' terminus of the M_2 dsRNA positive strand. This model is consistent with the data presented in Figure 8, beginning with the 5'-terminal G residue. (▷), T_1 ribonuclease cleavage, (▶), S_1 nuclease cleavage; solid lines represent strong cleavage; dashed lines represent weaker cleavage.

encoded on the M_2-1 fragment of M_2 dsRNA. Neither the genomic negative strand, nor denatured M_2-2 dsRNA programs the synthesis of detectable messenger-dependent polypeptides (data not shown). We also note that where we see M-p28, we also observe a slightly faster migrating "shadow" band. This may be due to premature termination, internal initiation, proteolysis, or an artifact of the migration of M-p28 through the gel.

A 1.05 kb RNA species (e.g., M_2-1) could theoretically encode a 38 kd polypeptide. Since denatured M_2-1 does not program the synthesis of the 38 kd species, but full-length m_2 transcript does, this 38 kd species (and perhaps) the

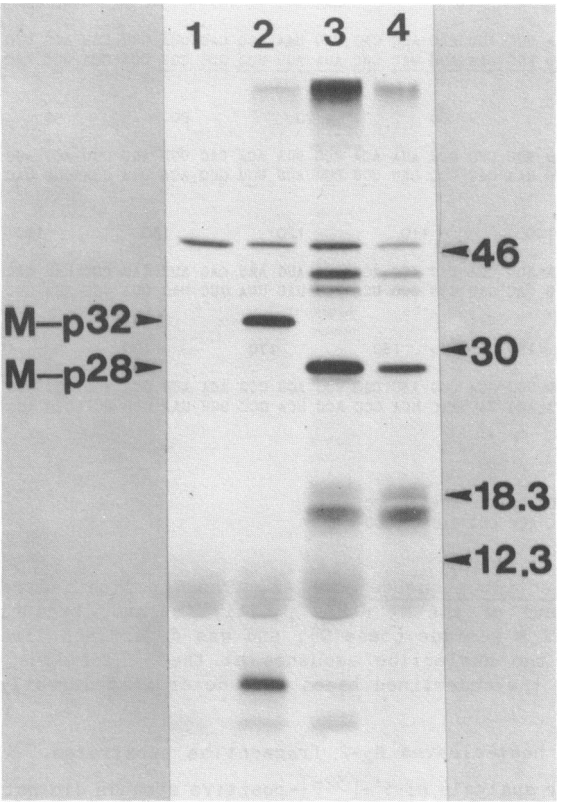


Figure 9. In vitro translation of M_2 dsRNA. M_1 dsRNA (30 ug/ml, lane 2) M_2 dsRNA (30 ug/ml, lane 3), or m_2 transcript synthesized in vitro (35 ug/ml, lane 4) were translated in vitro in a rabbit reticulocyte lysate system, in the presence of [^{35}S]-methionine, as described in Materials and Methods. Translation products were denatured and subjected to electrophoresis through a 12% polyacrylamide gel followed by fluorography. Numbers to the right indicate the positions of [^{14}C]-polypeptide standards (in kilodaltons). Lane 1, no RNA control.

40.5 kd species) may represent translation beyond the cleavage point of the M_2-1 fragment. We do not yet know if these species are immunologically related to each other or to the mature K_2 toxin. Consequently, we cannot entirely rule out the 38 kd or 40.5 kd species as potential toxin precursors. The smaller [^{35}S]-methionine-containing species observed may result from either premature termination and/or internal initiation, or proteolysis.

Terminal sequence of M_2-2

Terminal RNA sequence analysis of the M_2-2, the smaller heat-cleavage product derived from M_2 dsRNA, was performed as described for M_2-1, using both separated

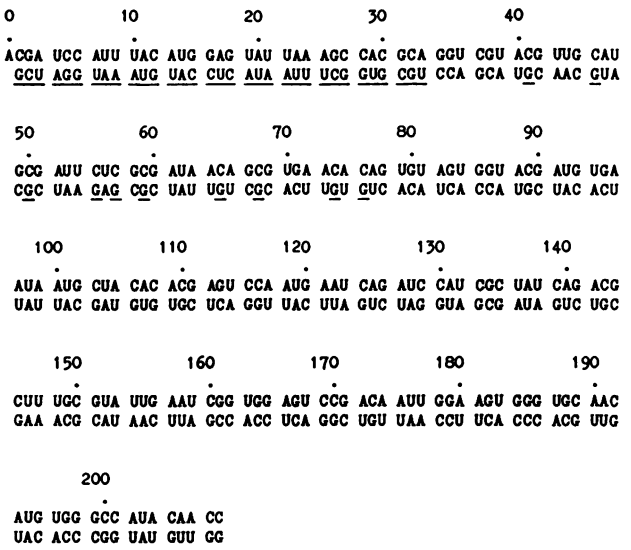


Figure 10. Nucleotide sequence of M_2-2. The top line indicates the sequence at the 3' terminus of the M_2 dsRNA positive strand, beginning with the 3'-terminal unpaired A residue (base 0), and was determined directly. The lower line represents the nucleotide sequence at the 5' terminus of the M_2 dsRNA negative strand; the underlined bases were determined directly.

strands and the heat-cleaved M_2-2 fragment as substrates. As we have noted, chemical sequence analysis of 3'-[^{32}P]-positive strands did not show pyrimidines which also reacted in the G-specific reaction. The results are summarized in Figure 10. As is the case with M_1-2 (29), the M_2-2 fragment contains no large open reading frames for as far as we have sequenced, which should be slightly more than half of the M_2-2 fragment.

There are some interesting homologies between the sequences of M_1-2 and M_2-2. This region, which includes the 3'-terminus of the positive strand, may contain sites required for encapsidation and, if replication occurs through full-length positive strands as an intermediate, for replicase recognition. M_2 dsRNA is competent for encapsidation within particles similar to those which contain M_1 dsRNA (4, Hannig, Leibowitz and Wickner, submitted). Interestingly, we find a conserved region of ten nucleotides, beginning at base 105 (numbered from the 3'-terminus of the positive strand) in both M_1 and M_2 dsRNAs, with the sequence 3'...ACACACGAGU...5'. Another region of homology is the sequence 3'...CUACACAC...5' beginning at base 103 from the 3' terminus of the M_2 dsRNA positive strand which, in M_1 dsRNA, is repeated beginning at position 1 at the 3'-terminus of the positive strand (the unpaired 3'-terminal A is numbered as position

0). Three seven-nucleotide homologies between M_1-2 and M_2-2 were present at positions 85 (M_1) and 202 (M_2), 135 (M_1) and 31 (M_2), and 213 (M_1) and 11 (M_2). Twelve homologies six nucleotides in length are also present in no apparent pattern.We do not find these sequences conserved at the corresponding terminus of L-A dsRNA isolated from a K_1 killer strain (9).

DISCUSSION

The M_1 and M_2 dsRNAs, which encode toxin and resistance functions in K_1 and K_2 Saccharomyces cerevisiae killer strains, respectively, possess extremely similar overall structures which may function in similar modes of expression and replication. These similarities may also be involved in the observed incompatability between these dsRNAs. Both contain an internal A,U-rich region which separates the molecule into a large (M-1) and a small (M-2) fragment. The positive polarity strand contains a polyadenylate-like tract in this A,U-rich region. In the case of M_1 dsRNA, the M_1-1 fragment contains the genetic information for toxin and resistance (17,22,23). A similar situation may exist in M_2 dsRNA, where the M_2-1 fragment programs the synthesis of a 28 kilodalton polypeptide in vitro. However, we cannot yet rule out certain other polypeptides synthesized in vitro from M_2-derived RNA species as K_2 toxin precursors. There is little homology between the N-terminal amino acids of the M_1 preprotoxin and the putative M_2 toxin precursor. However, there is striking similarity between these polypeptides in the presence of an extremely hydrophobic region and the predicted secondary structural features in this region. These N-terminal sequences are predicted to be capable of a membrane insertion function (50). There is some question, however, as to whether this region serves as a classical signal sequence (22) in the case of the M_1 preprotoxin. No such data are yet available on the M_2 toxin precursor.

In vitro transcription by the virion polymerase in K_1 and K_2 killers results in the production of full-length l and m transcripts. We have denoted the full-length positive polarity transcript of M_2 dsRNA as m_2. The 5' terminus of the positive strand of both M_1 and M_2 dsRNAs is capable of forming a stem-and-loop structure (19,25, and this report) containing the AUG which begins the open reading frame for the M_1 toxin precursor and, in the case of M_2, the AUG which may initiate the M_2 toxin precursor open reading frame. Both AUG triplets occur in favorable initiation contexts (44), both are in locally unstable regions of the secondary structure, and both are located near sequences that can theoretically base pair with the 3'-termini of yeast 18S and 5.8S ribosomal RNAs. In addition, some features of the secondary structure itself are conserved between M_1 and M_2 dsRNAs. As has been previously pointed out (49), the conservation of "A-richness" at the 5' terminus of the

positive strand of M_1 and M_2 dsRNAs may be important in transcriptase recognition. This A-richness presumably also contributes to the approximate 10-fold greater efficiency of the polynucleotide kinase reaction at this terminus in vitro. Our sequence at this terminus of M_2 completely agrees with a 20 base sequences previously reported (49).

The terminus of M_2 dsRNA represented in the M_2-2 fragment contains some interesting regions of homology with M_1-2, most notably the ten base pair conserved sequence beginning at position 105 from the 3'-terminus of the positive strand. A portion of this sequence is repeated at the extreme 3' terminus of the M_1 dsRNA positive strand. Whether these sequences, or other shorter sequences, are important in the functions of encapsidation and replicase recognition is not known. The conserved sequences at positions 105- 114 in M_1 and M_2 dsRNAs are part of runs of alternating purines and pyrimidines of 13 nucleotides on M_1 (positions 99-111) and 8 nucleotides on M_2 (positions 104-111). Although the configuration of these conserved regions is unknown, these sequences may form a Z-RNA double-helix under some conditions (53). The biological significance of such a conserved structure is unknown. However, similar sequences are not found at the 3' terminus of the L-A dsRNA positive strand (9). Our sequence at this terminus of M_2 agrees with a previously reported (49) 28 base sequence for the first 18 residues, but diverges at the nineteenth base. Secondary structural features present at the 3' terminus of the L-A dsRNA positive strand are absent from the 3' terminus of the M_1 dsRNA positive strand (9,25). The observed differences in primary and secondary structure between L-A and the M dsRNAs may, in part, underlie the differences in chromosomal genes required for the maintenance of these genomes. Of the 30 chromosomal genes known to be required for M_1 dsRNA maintenance (MAK genes), only MAK3, MAK10, and PET18 are known to be required for L-A dsRNA maintenance (2). Of the five MAK genes tested (MAK16, MAK10, MAK8, SPE2 and SPE10), all are required for M_2 dsRNA maintenance (2,16). In addition, M_2 requires MKT1 and MKT2 if [NEX] is present. It is not suprising that M_1 and M_2 dsRNAs require many of the same chromosomal genes for replication, since the observed incompatability between these two dsRNAs would seem to indicate competition between them (43). The similarity in the expresion of the killing phenotype of K_1 and K_2 killers is indicated by the requirement of KEX1 and KEX2 (killer expression) for expression of both the K_1^+ and K_2^+ phenotypes (R.B. Wickner, personal communication).

We have not yet characterized the in vivo transcripts of M_2 dsRNA. In the case of M_1 dsRNA, both full-length and subgenomic (1.2 kb) positive-polarity transcripts are observed in vivo (19,28). These polyadenylated transcripts presumably derive their poly(A) tract by transcription of the internal genomic A,U-rich region

(19,29). The subgenomic transcript is most likely translated in vivo to produce toxin and resistance factor. The full-length transcript may be translated in vivo, but may also serve as a replication intermediate. From the data presented here, we would anticipate a similar pattern of in vivo expression for M_2 dsRNA. The yeast M dsRNAs represent a class of viral dsRNAs which are widely divergent at the primary nucleotide and encoded amino acid sequence levels, but which share secondary structures and a few specific sequences which may be required for expression and replication.

ACKNOWLEDGEMENTS

We gratefully acknowledge Denise E. Georgopoulos for expert technical assistance in the preparation of virions from strain 1384, Dr. William T. McAllister for the use of computer facilities and Dr. Reed B. Wickner for personal communication of unpublished data. This investigation was supported in part by Public Health Service Grant GM-32413. E.M.H. is a predoctoral trainee supported by Public Health Service Institutional Award CA-09069 from the National Cancer Institute. M.J.L. is an Alexandrine and Alexander L. Sinsheimer Scholar.

REFERENCES

1. Bussey, H. (1981) Adv. Microb. Physiol. 22, 93-122.
2. Wickner, R.B. (1983) Arch. Biochem. Biophys. 222, 1-11.
3. Tipper, D.J. and Bostian, K.A. (1984). Microbiol. Rev. 48, 125-156.
4. Sommer, S.S. and Wickner, R.B. (1982) Cell 31, 429-441.
5. Hopper, J.E., Bostian, K.A., Rowe, L.B. and Tipper, D.J. (1977) J. Biol. Chem. 252, 9010-9017.
6. Bostian, K.A., Sturgeon, J.A. and Tipper, D.J. (1980) J. Bacteriol. 143, 463-470.
7. Bruenn, J., Bobek, L., Brennan, V. and Held, W. (1980) Nucleic Acids Res. 8, 2985-2997.
8. Sommer, S.S. and Wickner, R.B. (1982) J. Bacteriol. 150, 545-551.
9. Thiele, D.J., Hannig, E.M. and Leibowitz, M.J. (1984) Mol. Cell Biol. 4, 92-100.
10. Wickner, R.B. and Toh-E, A. (1982) Genetics 100, 159-174.
11. Wickner, R.B. (1980) Cell 21, 217-226.
12. Ridley, S.P., Sommer, S.S. and Wickner, R.B. (1984) Mol. Cell Biol. 4, 761-770.
13. Ball, S.G., Tirtiaux, C. and Wickner, R.B. (1984) Genetics 107, 199-217.
14. Welsh, J.D., Leibowitz, M.J. and Wickner, R.B. (1980) Nucleic Acids Res. 8, 2349-2363.
15. Welsh, J.D. and Leibowitz, M.J. (1980) Nucleic Acids Res. 8, 2365-2375.
16. Tyagi, A.K., Wickner, R.B., Tabor, C.W. and Tabor, H. (1984) Proc. Natl. Acad. Sci. USA, 81, 1149-1153.
17. Welsh, J.D. and Leibowitz, M.J. (1982) Proc. Natl. Acad. Sci. USA, 79, 786-789.
18. Fried, H.M. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 4224-4228.
19. Hannig, E.M., Thiele, D.J. and Leibowitz, M.J. (1984) Mol. Cell Biol. 4, 101-109.

20. Bostian, K.A., Hopper, J.E., Rogers, D.T. and Tipper, D.J. (1980) Cell 19, 403-414.
21. Bostian, K.A., Jayachandran, S. and Tipper, D.J. (1983) Cell 32, 169-180.
22. Bostian, K.A., Elliott, Q., Bussey, H., Burn, V., Smith, A. and Tipper, D.J. (1984) Cell 36, 741-751.
23. Lolle, S., Skipper, N., Bussey, H. and Thomas, D.Y. (1984) EMBO J. 3, 1383-1387.
24. Thiele, D.J., Wang, R. W. and Leibowitz, M.J. (1982) Nucleic Acids Res. 10, 1661-1677.
25. Thiele, D.J. and Leibowitz, M.J. (1982) Nucleic Acids. Res. 10, 6903-6918.
26. Leibowitz, M.J., Thiele, D.J. and Hannig, E.M. (1983) In D.H.L. Bishop and R.W. Compans (eds.), Double-stranded RNA viruses. Elsevier/North Holland Publishing Co., New York, pp. 457-462.
27. Skipper, N., Thomas, D.Y. and Lau, P.C.K. (1984) EMBO J. 3, 107-111.
28. Bostian, K.A., Burn, V.E., Jayachandran, S. and Tipper, D.J. (1983) Nucleic Acids Res. 11, 1077-1097.
29. Thiele, D.J., Hannig, E.M. and Leibowitz, M.J. (1984) Virology 137, 20-31.
30. Naumov, G.I. and Naumova, T.I. (1973) Genetika 9, 140-145.
31. Wickner, R.B. (1983) Mol. Cell Biol. 3, 654-661.
32. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA, 69, 1408-1412.
33. Laemmli, U.K. (1970) Nature (London) 227, 680-685.
34. Bonner, W.M. and Laskey, R.A. (1974) Eur. J. Biochem. 46, 83-88.
35. Peattie, D. (1979) Proc. Natl. Acad. Sci. USA, 76, 1760-1764.
36. Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) Nucleic Acids Res. 4, 2527-2538.
37. Simoncsits, A., Brownlee, G.G., Brown, R.S., Rubin, J.R. and Guilley, H. (1977) Nature (London) 269, 833-836.
38. Volckaert, G. and Fiers, W. (1977) Anal. Biochem. 83, 222-227.
39. Pavlakis, G.N., Lockard, R.E., Vamvakopoulos, N., Rieser, L., RajBhandary, U.L. and Vournakis, J.N. (1980) Cell 19, 91-102.
40. Gordon, K.H.J. and Symons, R.H. (1983) Nucleic Acids Res. 11, 947-960.
41. Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) Biochemistry 16, 4743-4751.
42. Bailey, J.N. and McAllister, W.T. (1980) Nucleic Acids Res. 8, 5071-5088.
43. Sommer, S.S. and Wickner, R.B. (1984) Mol. Cell. Biol. 4, 1747-1753.
44. Kozak, M. (1984) Nucleic Acids Res. 12, 857-872.
45. Hitzeman, R.A., Hagie, F.E., Levine, H.L., Goeddel, D.V., Ammerer, G. and Hall, B.D. (1981) Nature (London) 293, 717-722.
46. Rubstov, P.M., Musakhanov, M.M., Zakharyev, V.M., Krayer, A.S., Skryabin, K.G. and Bayev, A.A. (1980) Nucleic Acids Res. 8, 5779-5794.
47. Rubin, G.M. (1973) J. Biol. Chem. 248, 3860-3875.
48. Tinoco, Jr., I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
49. Brennan, V.E., Field, L., Cizdziel, P. and Bruenn, J.A. (1981) Nucleic Acids Res. 9, 4007-4021.
50. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
51. Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120.
52. Perlman, D. and Halvorson, H.O. (1983) J. Mol Biol. 167, 391-409.
53. Hall, K., Cruz, P., Tinoco, Jr., I., Jovin, T.M. and van de Sande, J.H. (1984) Nature (London) 311, 584-586.