Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

VIROLOGY 191, 992–995 (1992)

Rice Dwarf Phytoreovirus Segment S12 Transcript Is Tricistronic *in Vitro*

NOBUHIRO SUZUKI,[1] MIKIKO SUGAWARA, AND TOMONOBU KUSANO

Laboratory of Plant Genetic Engineering, Biotechnology Institute, Akita Prefectural College of Agriculture, Ohgata, Akita 010-04, Japan

Received June 24, 1992; accepted August 13, 1992

Sequence analysis revealed that rice dwarf phytoreovirus segment S12 is 1066 nucleotides long with a small out-of-phase, overlapping open reading frame (ORF) as well as a major ORF. The large ORF (positions 42 to 980) encodes 312 amino acids, while the small one (bases 313 to 591) encodes 92 amino acids with an additional in-frame AUG codon (positions 337–339) 24 nucleotides downstream from the first one. Transcripts from a full-length cDNA directed the *in vitro* synthesis of three polypeptides of 33 (considered to be translated from the long ORF), 8, and 7 kDa. Alteration of each of the two ATG codons on the small ORF demonstrated their involvement in the generation of the 8- and 7-kDa polypeptides. Although it is still unknown whether these proteins are expressed *in vivo,* the small ORF is shown to be conserved in S9s of two other members of the genus *Phytoreovirus,* rice gall dwarf virus and wound tumor virus, suggesting its common, important function. © 1992 Academic Press, Inc.

Eukaryotic mRNAs are generally monocistronic, with 5′-proximal AUGs which are used as start sites for translation. However, examples of mRNAs which initiate at more than one AUG have been described. These mRNAs are translated into two (or more) completely different proteins or long and short protein isoforms (*1*). Most cases are limited to mRNAs of viruses including influenza B virus (*2*), infectious bronchitis coronavirus (*3*), reovirus (*4, 5*), rotavirus (*6*), barley yellow dwarf luteovirus (*7*), turnip yellow mosaic tymovirus (*8*) and cucumber necrosis tombusvirus (*9*). Rice dwarf phytoreovirus (RDV) (*10, 11*), a member of the family *Reoviridae,* has a genome composed of 12-segmented double-stranded (ds) RNA (S1–S12). Each segment reported so far has been shown to have a single long open reading frame (ORF) (*12–21*). However, during the course of sequence analysis of RDV S12, we have found that the segment has an out-of-frame, overlapping small ORF as well as a large ORF. In the current study, we determined the complete nucleotide sequence of RDV S12 and showed that transcripts from a full-length cDNA of S12 specified the large ORF-encoded polypeptide and long and short versions of the small ORF-coded proteins in an *in vitro* wheat germ translation system. Furthermore, we found that the corresponding small ORF is conserved in S9s of two other members of the genus *Phytoreovirus,* rice gall dwarf virus (RGDV) (*22*) and wound tumor virus (WTV) (*23*).

Sequence data from this article have been deposited with the EMBL/GenBank/DDBJ Data Libraries under Accession No. D90200.

[1] To whom reprint requests should be addressed.

From the cDNA library of the RDV mRNAs constructed previously (*16*), two S12 cDNA clones (p12RD1, p12RD2) were newly selected. p12RD2 was recloned into the pUC18 polylinker site (*24*) in the orientation opposite to that of the original plasmids. Deletion plasmids, at intervals of about 200 bp, were produced from the original and recloned cDNA inserts by digestion from one end with exonuclease III (*25*). The deleted cDNAs were sequenced by the dideoxynucleotide chain-termination method (*26*) using a Sequenase Version 2.0 DNA sequencing kit (United States Biochemical Corp. Cleveland, OH), in which 7-deaza-dGTP was utilized instead of dGTP (*27*). The entire sequence was analyzed in both directions.

It was ascertained as in the case of S4 (*18*) that p12RD2 covered the 3′ terminus of the RDV S12 transcript. p12RD1 and p12RD2 covered bases 24–1064 and bases 39–1066. Bases 1–23 were determined by sequencing S12 mRNA, in which the oligodeoxynucleotides complementary to bases 70–88 was used as a primer in the presence of dideoxynucleoside triphosphate (*28*). RDV S12 has 1066 base pairs and possesses large and small ORFs in different frames extending for 936 nucleotides from bases 42–977 and 276 nucleotides from bases 313–588, respectively. The nucleotide and deduced amino acid sequences are shown in Fig. 1. The long ORF encodes 312 amino acids with a M_r of 33,919, while the other encodes 92 amino acids with a M_r of 10,551. The small ORF additionally contains a downstream in-frame ATG codon at positions 337–339.

Wild type, deletion mutant, and site-directed mutant cDNAs were made and used for *in vitro* transcription

0042-6822/92 $5.00
Copyright © 1992 by Academic Press, Inc.
All rights of reproduction in any form reserved.

```
                              GGTAAATTGAGCAGTATTTCAACATTGTTGTTGGAGTTATA      41

ATGTTCAAGAGCGGGTCCGGTTCTTTGAAGCGAAGCGGGTCCATTTCAAGTGTTAAATCA     101
M  F  K  S  G  S  G  S  L  K  R  S  G  S  I  S  S  V  K  S        20

TTTTCTGGCGATAGTGAAAAAGGACTACCACCTATTTCCCGAGGTAGTGTGTCCATTGCT     161
F  S  G  D  S  E  K  G  L  P  P  I  S  R  G  S  V  S  I  A        40

AGTCAGAACTCTGAACCATTGATTGTCCCCGCAAGTAGTTCTTCTTTTGCGGCGACTTCT     221
S  Q  N  S  E  P  L  I  V  P  A  S  S  S  S  F  A  A  T  S        60

GACTTTGTACCTGAAAAGACAAAGTCCGAAGGGAATTTGAAAGATAAGTCTTCCGTCATA     281
D  F  V  P  E  K  T  K  S  E  G  N  L  K  D  K  S  S  V  I        80

                                       M  L  I  L  I  R  T  L  M  A      10
ACCGGAAACTTTGGGTCCTCTGGACCAATTAATGCTCATACTAATCAGAACGCTGATGGC     341
T  G  N  F  G  S  S  G  P  I  N  A  H  T  N  Q  N  A  D  G       100

 I  D  W  L  K  I  F  C  L  R  N  R  P  K  D  E  D  Q  A  H      30
GATCGATTGGTTGAAAATCTTTTGCTTAAGGAATCGTCCAAAGGACGAGGATCAGGCACA     401
D  R  L  V  E  N  L  L  L  K  E  S  S  K  G  R  G  S  G  T       120

 P  M  L  D  I  L  P  Q  T  L  D  S  V  K  K  S  S  S  R  F      50
TCCGATGCTAGACATACTGCCACAGACTCTCGACTCAGTCAAGAAGTCAAGCAGTCGTTT     461
S  D  A  R  H  T  A  T  D  S  R  L  S  Q  E  V  K  Q  S  F       140

 P  R  K  M  L  A  A  T  I  S  I  L  E  E  E  V  T  E  L  V      70
TCCGAGGAAAATGCTGGCGGCAACGATCTCAATACTGGAAGAGGAAGTCACGGAACTGGT     521
S  E  E  N  A  G  G  N  D  L  N  T  G  R  G  S  H  G  T  G       160

 T  E  L  N  N  T  T  N  L  T  A  K  K  E  C  P  R  I  T  N      90
GACGGAGTTGAACAACACTACAAATTTGACTGCGAAGAAGGAATGTCCGCGTATCACAAA     581
D  G  V  E  Q  H  Y  K  F  D  C  E  E  G  M  S  A  Y  H  K       180

 A  L  *     92
CGCGTTGTAGATACATTCTTTAAATATTTTGAATATTCGGCTGAAGATGGGCATTCCACC     641
R  V  V  D  T  F  F  K  Y  F  E  Y  S  A  E  D  G  H  S  T       200

CTATATTCGGACGTTATGTTTTTATCTGGTCATGGTGATCTTGGTCTGCTAGTTATGAGT     701
L  Y  S  D  V  M  F  L  S  G  H  G  D  L  G  L  L  V  M  S       220

AGATATCAAGAACTCATGACCTTACGGGTGCGAAGTGCTATTTATGGCATATTTTGTTAC     761
R  Y  Q  E  L  M  T  L  R  V  R  S  A  I  Y  G  I  F  C  Y       240

CTACAAGCTCTAACCGCCTACCTAACGTATTTTGACGCTAAAGTGGGTCAAGCTATTATG     821
L  Q  A  L  T  A  Y  L  T  Y  F  D  A  K  V  G  Q  A  I  M       260

TTGGATGAGGAGTTGGAGAAATATGAAATCCGACTCGATGTCGCGCAAGACGACGATCCA     881
L  D  E  E  L  E  K  Y  E  I  R  L  D  V  A  Q  D  D  D  P       280

ATTGTATTTCAAATCACGACGGGTGTATTCACATCGGGAGTAGCTCACGATCTGCGTAAA     941
I  V  F  Q  I  T  T  G  V  F  T  S  G  V  A  H  D  L  R  K       300

CTTACACAGATACTCGAGGCCTTTTCTCTTGAACGGTGATATCCAGTCTGGGAGTAATGA    1001
L  T  Q  I  L  E  A  F  S  L  E  R  *     312

ACCCACGCTGGGGACTCAGTTATATCTCTTCGGGAGTGATATACCGAGTACTGCTCATAA    1061

CTGAT     1066
```

Fig. 1. Nucleotide and presumed amino acid sequences of the mRNA-sense cDNA of RDV segment S12. The nucleotides (bases 70–88) complementary to the synthesized primer used in direct sequencing of S12 mRNAs are underlined. Deduced amino acid sequences for the large ORF (bases 42–980) and small ORF (bases 313–591) are shown below and above the nucleotide sequence, respectively. Stop codons are indicated by asterisks.

and translation in order to define S12 translational products. Detailed information on cloning of these clones is described in the legend of Fig. 2. Synthetic transcripts were obtained using T3 or T7 RNA polymerase from the cDNAs linearized by *Kpn*I or *Sac*II according to the manufacturer's instructions (Stratagene). Three micrograms of the obtained transcript was subjected to *in vitro* translation in wheat germ extracts, and the prod

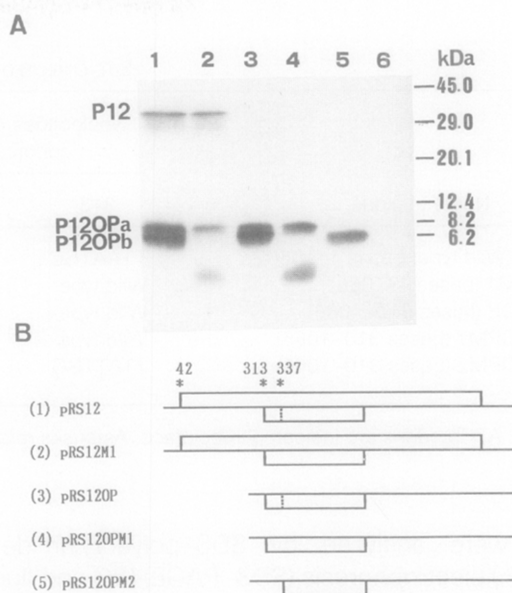

Fig. 2. *In vitro* translation analysis using synthetic RDV S12 transcripts. (A) Fluorogram of SDS–PAGE of *in vitro* translation products specified by transcripts from a full-length cDNA (pRS12) and its deletion and site-directed mutants. Transcripts used were from pRS12 (lane 1), pRS12M1 (lane 2), pRS12OP (lane 3), pRS12OPM1 (lane 4), pRS12OPM2 (lane 5), and endogenous wheat germ mRNAs (lane 6). The migration positions of protein size markers (ovalbumin, 45 kDa; carbonic anhydrase, 29 kDa; soybean trypsin inhibitor, 20.1 kDa; cytochrome *c*, 12.4 kDa; myoglobin fragment III, 8.2 kDa; myoglobin fragment II, 6.2 kDa) are indicated on the left. A full-length cDNA (pRS12) was synthesized by the polymerase chain reaction (PCR) (*29*) using p12RD1 as a template, in which 5′GCCGCGGTAAATTGAGCAGTATTTCAACATTGTTGTTGGAGTTATAATGTTCAAGAGCGGGTCCGG3′ (PI) (underlined sequence is the *Sac*II site added to the terminal sequence of S12) and 5′CGGTACCATCAGTTATGAGCAGTACTC3′ (PII) (sequence underlined is the *Kpn*I site added to the S12 terminal end) served to prime, and then cloned into pBluescriptII SK(+) (product literature from Stratagene Cloning System, La Jolla, CA). A deletion mutant (pRS12OP) covering bases 310–1066 was synthesized by PCR using 5′CCCCGCGGTTAATGCTCATACTAATCA3′ (PIII) corresponding to bases 310–328 (sequence underlined is the *Sac*II site) and PII as primers. As a site-directed mutant (pRS12M1), a *Cla*I–*Kpn*I fragment of pRS12 (bases 345–1066) and a *Cla*I-digested cDNA (p12RD3, bases 1–344) with a point mutation at the second ATG codon of the small ORF were ligated. p12RD3 was made by PCR using PI and 5′CCAATCGATCGCTATCAGCGT3′ complementary to bases 331–351 (base T marked by an asterisk was introduced in place of C, and the underlined sequence is the *Cla*I site). Two other site-directed mutants (pRS12OPM1, pRS12OPM2) covering bases 310–1066 were also constructed by using the PCR technique. The templates and primers used were pRS12M1, PII, and PIII for pRS12OPM1, and pRS12OP, PII and 5′CACCGCGGTTATTGGTCATACT3′ (underlined sequence is the *Sac*II site and bases indicated by asterisks are point mutations) for pRS12OPM2. These mutants are identical with pRS12 except for the nucleotide changes shown in Table 1. (B) Schematic representation of genetic organization of the cDNAs used in Fig. 2A. The start sites for the small and large ORFs are indicated by asterisks and their positions are also shown on the top. The second in-frame ATG triplet of the small ORF is indicated by a dashed line.

TABLE 1

SITE-DIRECTED AND DELETION MUTANTS OF pRS12

Name of clone	Nucleotides around the ATG codon at position		Production of polypeptides of		
	313	337	33 kDa (P12)	8 kDa (P12OPa)	7 kDa (P12OPb)
pRS12 (Wild type, bases 1–1066)	TTA**ATG**C	CTG**ATG**G	+	+	+
pRS12M1 (bases 1–1066)	Wild type	CTGATA*G	+	+	−
pRS12OP (bases 310–1066)	Wild type	Wild type	−	+	+
pRS12OPM1 (bases 310–1066)	Wild type	CTGATA*G	−	+	−
pRS12OPM2 (bases 310–1066)	TTAT*TGG*	Wild type	−	−	+

Note. ATG codons are indicated in boldface. Asterisks refer to nucleotide changes from the wild type.

ucts were analyzed by SDS–polyacrylamide gel (17.5%) electrophoresis (SDS–PAGE) (*30*) and fluorography according to the method of Shirako and Ehara (*31*). As a result, the transcripts from a wild-type full-length cDNA (pRS12) directed three species of polypeptides with M_r values of 33 kDa (P12), 8 kDa (P12OPa), and 7 kDa (P12OPb), respectively (Fig. 2, lane 1). The transcripts from pRS12M1, in which the second ATG codon of the small ORF was absent, programmed the synthesis of P12 and P12OPa (Fig. 2, lane 2). The transcripts from pRS12OP, which does not contain the large ORF, specified both P12OPa and P12OPb, but not P12 (Fig. 2, lane 3). The pRS12OPM1 transcript directed only the synthesis of P12OPa, while pRS12OPM2 specified P12OPb alone (Fig. 2, lanes 4 and 5). pRS12OPM1 and pRS12OPM2 lacked the first and second ATG codons of the small ORF, respectively. No polypeptides were detected from the minus-sense transcripts from these cDNAs (data not shown). These results are summarized in Table 1, demonstrating that P12, P12OPa and P12OPb were translated from the large ORF, and the first and second AUG codons of the small ORF, respectively.

The genus *Phytoreovirus* has three members, namely, WTV, RGDV, and RDV (*11*). The previous sequence analyses (*13, 16, 18, 23, 32*) showed the corresponding segment assignment of WTV and RDV, and suggested that RDV S12 and WTV S9 remain to be assigned on the basis of size similarity. However, no amino acid sequence homology with a score of more than 33 (*33*) was found between RDV P12 and the WTV S9-encoded nonstructural protein (Pns10, 345 amino acids). With regard to RGDV, RGDV S9 has been shown to have a 5′-noncoding sequence virtually identical to that of WTV S9 (*22*), although the RGDV S9-encoded polypeptide (323 amino acids) has no significant sequence homology with WTV Pns10 or any RDV-coded proteins. Surprisingly, however, careful inspection revealed that both WTV S9 and RGDV S9 also encode small polypeptides (P9OPs) which have significant sequence homologies, scores of which are more than 100, with RDV P12OP. The small ORFs of WTV S9 and RGDV S9 were previously unrecognized. The RGDV S9 small ORF encoding 76 amino acids (P9OP) starts with an ATG codon at bases 234–236 and ends with a termination codon at bases 462–464, while the WTV S9 ORF extends from bases 240–482, which encodes 80 amino acids (P9OP). The sequence alignment and identity among the three viruses are shown in Fig. 3. The amino acid sequence similarities between RDV P12OPa and RGDV P9OP, between RDV P12OPa and WTV P9OP, and between RGDV P9OP

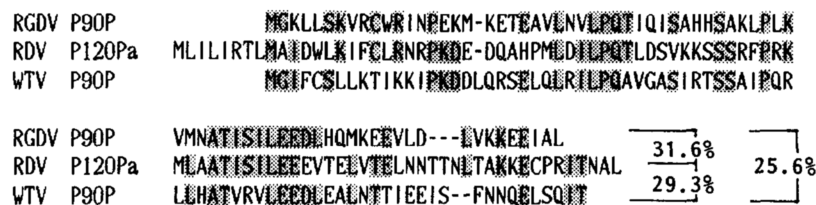

FIG. 3. Comparison of deduced amino acid sequences of polypeptides encoded by small open reading frames of RGDV S9, RDV S12 and WTV S9. Percentage identity in the optimal pairwise alignments including gaps are indicated to the left of the second column. Identical amino acid residues, which are the same for at least two viruses, are shaded. Amino acid sequence of RDV S12 was cited from this study, that of RGDV S9 from Koganezawa *et al.* (*22*) and that of WTV S9 from Anzola *et al.* (*23*).

and WTV P9OP are 31.6, 29.3, and 25.6%, respectively. Two regions (residues 33 to 39 and residues 57 to 64 on RDV P12OPa) are well conserved. Thus the three segments seem to have the same ancestral origin, and the small ORF might have been more conserved than the large one during the course of evolution.

No significant sequence homology was found between either RDV P12 or RDV P12OPa and proteins from the National Biomedical Research Foundation protein sequence library (Release 31), even between RDV P12OPa and the out-of-phase, small ORF-coded proteins of other reoviruses and plant viruses.

In the present study, RDV S12 proved to specify the *in vitro* expression of three polypeptide products. P12 (33 kDa) was assumed to be translated from the long ORF, while P12OPa and P12OPb were considered to be isoforms from the same reading frame. Although the mechanism by which RDV S12 is tricistronic remains unknown, leaky scanning is a distinct possibility, since the first AUGs of the large and small ORFs deviate from the consensus sequence (*34*) in either or both positions −3 and +4, and no frame contains an AUG triplet between the first AUG codons of the two ORFs. The conservation of the internal small ORF in the other phytoreoviruses, WTV and RGDV, strongly suggests that the small ORF-encoded protein may be expressed *in vivo* and play an important role during their life cycle.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Marilyn Kozak for useful information on tricistronic mRNAs, Dr. Kripamoy Aguan for fruitful discussion, and to Ms. Kouko Futada for typing the manuscript. This work was partly supported by the Nissan Science Foundation.

REFERENCES

1. KOZAK, M., *J. Cell Biol.* **115,** 887–903 (1991).
2. WILLIAMS, M. A., and LAMB, R. A., *J. Viol.* **63,** 28–35 (1989).
3. LIU, D. X., CAVANAGH, D., GREEN, P., and INGLIS, S. C., *Virology* **184,** 531–544 (1991).
4. ERNST, H., and SHATKIN, A., *Proc. Natl. Acad. Sci. USA* **82,** 48–52 (1985).
5. SARKER, G., PELLETIER, J., BASSEL-DUBY, R., JAYASURIYA, A., FIELDS, B. N., and SONENBERG, N., *J. Virol.* **54,** 720–725 (1985).
6. MATTION, N. M., MITCHELL, D. B., BOTH, G. W., and ESTES, M. K., *Virology* **181,** 295–304 (1991).
7. DINESH-KUMAR, S. P., BRAUT, V., and MILLER, W. A., *Virology* **187,** 711–722 (1992).
8. WEILAND, J. J., and DREHER, T. W., *Nucleic Acids Res.* **17,** 4675–4687 (1989).
9. ROCHON, D. M., and JOHNSTON, J. C., *Virology* **181,** 656–665 (1991).
10. IIDA, T. T., SHINKAI, A., and KIMURA, I., "CMI/AAB Descriptions of Plant Viruses," No. 102 (1972).
11. BOCCARDO, G., and MILNE, R. G., "CMI/AAB Descriptions of Plant Viruses," No. 294 (1984).
12. FUKUMOTO, F., OMURA, T., and MINOBE, Y., *Arch. Virol.* **107,** 135–139 (1989).
13. NAKASHIMA, K., KAKUTANI, T., and MINOBE, Y., *J. Gen. Virol.* **71,** 725–729 (1990).
14. OMURA, T., ISHIKAWA, K., HIRANO, H., UGAKI, M., MINOBE, Y., TSUCHIZAKI, T., and KATO, H., *J. Gen. Virol.* **70,** 2759–2764 (1989).
15. OMURA, T., MINOBE, Y., and TSUCHIZAKI, T., *J. Gen. Virol.* **69,** 227–231 (1988).
16. SUZUKI, N., HARADA, M., and KUSANO, T., *J. Gen. Virol.* **72,** 2233–2238 (1991).
17. SUZUKI, N., TANIMURA, M., WATANABE, Y., KUSANO, T., KITAGAWA, Y., SUDA, N., KUDO, H., UYEDA, I., and SHIKATA, E., *Virology* **190,** 240–247.
18. SUZUKI, N., WATANABE, Y., KUSANO, T., and KITAGAWA, Y., *Virology* **179,** 446–454 (1990).
19. SUZUKI, N., WATANABE, Y., KUSANO, T., and KITAGAWA, Y., *Virology* **79,** 455–459 (1990).
20. UYEDA, I., KUDO, H., TAKAHASHI, T., SANO, T., OHSHIMA, K., MATSUMURA, T., and SHIKATA, E., *J. Gen. Virol.* **70,** 1297–1300 (1989).
21. UYEDA, I., MATSUMURA, T., SANO, T., OHSHIMA, K., and SHIKATA, E., *Proc. Jpn. Acad. Series B* **63,** 227–230 (1987).
22. KOGANEZAWA, H., HIBINO, H., MOTOYOSHI, F., KATO, H., NODA, H., ISHIKAWA, K., and OMURA, T., *J. Gen. Virol.* **71,** 1861–1863 (1990).
23. ANZOLA, J. V., DALL, D. J., XU, Z., and NUSS, D. L., *Virology* **171,** 222–228 (1989).
24. NORRANDER, J., KEMPE, T., and MESSING, J., *Gene* **26,** 101–106 (1983).
25. YANISCH-PERRON, C., VIEIRA, J., and MESSING, J., *Gene* **33,** 103–119 (1985).
26. SANGER, F., NICKLEN, S., and COULSON, A. R., *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467 (1977).
27. MIZUSAWA, S., NISHIMURA, S., and SEELA, F., *Nucleic Acids Res.* **14,** 1319–1324 (1986).
28. DEBORDE, D. C., NAEVE, C. W., HERLOCHER, M. L., and MAASSAB, H. F., *Anal. Biochem.* **157,** 275–282 (1986).
29. SAIKI, R. K., GELFAND, D. H., STOFFE, S., SHARF, S. J., HIGUCHI, R., HORN, G. T., MULLIS, K. B., and ERLICH, H. A., *Science* **239,** 487–491 (1988).
30. LAEMMLI, U. K., *Nature (London)* **227,** 680–685 (1987).
31. SHIRAKO, Y., and EHARA, Y., *J. Gen. Viol.* **67,** 1237–1245 (1986).
32. DALL, D. J., ANZOLA, J. V., XU, Z., and NUSS, D., *Nucleic Acids Res.* **17,** 3599 (1989).
33. LIPMAN, D. J., and PEARSON, W. R., *Science* **227,** 1435–1441 (1985).
34. KOZAK, M., *Nucleic Acids Res.* **15,** 8125–8132 (1987).