JOURNAL OF VIROLOGY, Sept. 2005, p. 12095–12099 Vol. 79, No. 18
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.18.12095–12099.2005

Identification and Characterization of Human Cytomegalovirus-Encoded MicroRNAs

Finn Grey,[1]* Andy Antoniewicz,[2] Edwards Allen,[4] Julie Saugstad,[3] Andy McShea,[2] James C. Carrington,[4] and Jay Nelson[1]

Department of Molecular Microbiology and Immunology, Oregon Health Sciences University, Portland, Oregon 97201[1]; Combimatrix Corporation, 6500 Harbour Heights Parkway, Mukilteo, Washington 98275[2]; Robert S. Dow Neurobiology Laboratories, Legacy Research, Portland, Oregon 97232[3]; and Center for Gene Research and Biotechnology and Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331[4]

Received 29 April 2005/Accepted 20 June 2005

MicroRNAs (miRNAs) are an extensive class of noncoding genes that regulate gene expression through posttranscriptional repression. Given the potential for large viral genomes to encode these transcripts, we examined the human cytomegalovirus AD169 genome for miRNAs using a bioinformatics approach. We identified 406 potential stem-loops, of which 110 were conserved between chimpanzee cytomegalovirus and several strains of human cytomegalovirus. Of these conserved stem-loops, 13 exhibited a significant score using the MiRscan algorithm. Examination of total RNA from human cytomegalovirus-infected cells demonstrated that 5 of the 13 predicted miRNAs were expressed during infection. These studies demonstrate that human cytomegalovirus encodes multiple conserved miRNAs and suggest that human cytomegalovirus may utilize an miRNA strategy to regulate cellular and viral gene function.

MicroRNAs (miRNAs) are a large class of noncoding RNAs involved in posttranscriptional regulation through RNA interference. A number of recent studies have identified virally encoded miRNAs using either biochemical cloning strategies, bioinformatics, or a combination of the two approaches (1, 2, 4, 7–9). In this study an alternative bioinformatics approach based on comparative conservation between predicted stem-loop sequences of human cytomegalovirus (HCMV) and chimpanzee cytomegalovirus (CCMV) was used to predict candidate miRNAs. Expression of predicted miRNAs during HCMV infection was then assessed by Northern blot analysis. Our bioinformatics approach utilized a computer algorithm called Stem-loop Finder (SLF; Combimatrix) to predict potential RNA transcripts from the HCMV genome that could form stem-loop secondary structures. The algorithm uses free energy calculations to determine the theoretical stability of the base pairing within the stem region, including pairing between G · U bases, while maintaining a maximum and minimum length between the complementary base pairing to determine loop size. A scoring matrix then weights beneficial or detrimental folding structures and attributes a cumulative score to each potential stem-loop sequence. Analysis of the HCMV genome using SLF identified 406 potential stem-loop sequences. Previously identified miRNAs are often extensively conserved between many different species (6). Consequently, we hypothesized that functionally important miRNAs expressed by HCMV would be conserved between HCMV and closely related viruses, such as CCMV. The sequence of each of the 406 candidate SLF-derived HCMV stem-loop transcripts was compared with the CCMV genome for potential homology. A minimum score of 60% homology with CCMV was used to select stem-loop sequences for further analysis. Our preliminary studies determined that this level of homology was sufficiently stringent to identify significant sequence conservation without exclusion of any potential miRNAs. Of the 406 sequences analyzed, 110 potential stem-loop sequences scored higher than 60% homology.

The 110 HCMV stem-loop sequences selected using the criteria (detailed above) and the corresponding homologous CCMV sequences were then analyzed using a bioinformatics program (MiRscan) (5) to predict which of the conserved stem-loop sequences had a high probability of encoding genuine miRNA transcripts. The MiRscan program compares two sequences and provides a score based on a number of aspects, such as the ability to form a stem-loop, symmetry of bulge loops, and conservation of the predicted miRNA sequence within the stem-loop. Specific sequence bias within the 5′ region of the miRNA, such as a propensity for the first base of the miRNA to be a uracil, is also a criterion for the program. The program then predicts a candidate miRNA that would be generated from the pre-miRNA stem-loop structure. In a similar previous study, stem-loop sequences that scored higher than 10 using the MiRscan algorithm were experimentally tested (5). The 110 HCMV stem-loop sequences selected using the criteria mentioned above and the corresponding homologous CCMV sequences were analyzed using the MiRscan program. Of the 110 sequences analyzed, 13 scored higher than 10 using the MiRscan algorithm. The 13 candidate miRNAs are shown in Table 1, with the predicted secondary structures shown in Fig. 1. To determine whether the predicted miRNAs were expressed by HCMV, Northern blot analysis of total RNA from infected cells was performed using end-labeled oligonucleotide probes. In each case probes complementary to both the full 5′ stem sequence and 3′ stem sequence, as well as probes for the predicted miRNA sequence, were used. Follow-

* Corresponding author. Mailing address: Department of Molecular Microbiology and Immunology, Oregon Health Sciences University, Portland, OR 97201. Phone: (503) 494-2434. Fax: (503) 494-6862. E-mail: greyf@ohsu.edu.

TABLE 1. Predicted miRNA sequences of HCMV[a]

Stem-loop	Conservation score vs CCMV	MiRscan score	Nucleotide coordinates of stem-loop	Relative genome position	Predicted miR
UL22-1	64	10.27	C 27639 to 27716	Between UL22-UL23	UCACGGGAAGGCUAGUUAGAC
UL31-1	80	11.80	C 39176 to 39258	A/S UL31	CGGCAUGUUGCGCGCCGUGAU
UL36-1	87	12.85	C 49502 to 49563	UL36 intron	AGACACCUGGAAAGAGGACGU
UL53-1	90	12.24	C 77054 to 77134	A/S UL53	UGCGCGAGACCUGCUCGUUGC
UL54-1	82	12.57	C 79184 to 79276	UL54	UGCGCGUCUCGGUGCUCUCGG
UL70-1	84	10.28	104018 to 104083	A/S UL70	UGCGUCUCGGCCUCGUCCAGA
UL102-1	87	13.39	148054 to 148113	UL102	UGGCCAUGUCGUUUCGCGUCG
UL102-2	82	12.95	C 148703 to 148767	A/S UL102	UGGCGUCGUCGCUCGGCGGGU
UL111a-1	93	14.64	161368 to 161436	Between UL111a-UL112/113	UGACGUUGUUUGUGGGUGUUG
US4-1	73	12.86	196080 to 196163	Partial US4	CGACAUGGACGUGCAGGGGGA
US5-1	62	14.77	196991 to 197056	Between US6-US7	UGACAAGCCUGACGAGAGCGU
US5-2	83	10.49	197120 to 197184	Between US6-US7	UGAUAGGUGUGACGAUGUCUU
US29-1	89	15.43	221396 to 221479	US29	UUGGAUGUGCUCGGACCGUGA

[a] For each of the 13 identified candidate miRNA stem-loop sequences, the nucleotide coordinates and genome positions relative to the annotated ORFs of HCMV (AD169) are shown. The percentage conservation compared to CCMV and the overall MiRscan score are also indicated along with the predicted miRNA sequence. C, complementary strand; A/S, antisense to ORF shown.

ing infection of primary fibroblast cells with HCMV AD169, total RNA was harvested at 2, 8, 24, 48, and 72 h postinfection (p.i.). Total RNA from uninfected fibroblast cells was included as a negative control. Four of the predicted miRNAs, UL36-1, US4-1, US5-1, and US5-2, were detected in infected RNA samples, with an additional miRNA, UL70-1, detected using a probe to the complementary strand of the stem-loop (Fig. 2). UL36-1 expression was first detected 24 h p.i. and continued to accumulate over time, with peak levels being detected 72 h p.i. A band of approximately 80 bases in length was also detected as early as 8 h p.i. and was present throughout the remainder of the time course. The pre-miRNA stem-loop transcript can often be detected in addition to the miRNA species as a band 60 to 80 bases in length. Therefore, the ~80-base species detected in the UL36-1 blot correlates with the pre-miRNA of UL36-1. The probe for US4-1 detected a band of approximately 24 to 25 bases in length, slightly larger than the other viral miRNAs. Previous cloning studies have determined that the 3′ end of miRNAs often varies by one or two bases (5), which was consistent with our detection of a smear of smaller RNA species below the major US4-1 band that resolved into at least two individual bands upon further analysis (Fig. 3). Expression of US5-1 and US5-2 was also first detected 24 h p.i. and, like UL36-1, continued to accumulate throughout the time course. In previous studies, the complementary strand of the miRNA duplex formed from the stem-loop precursor was

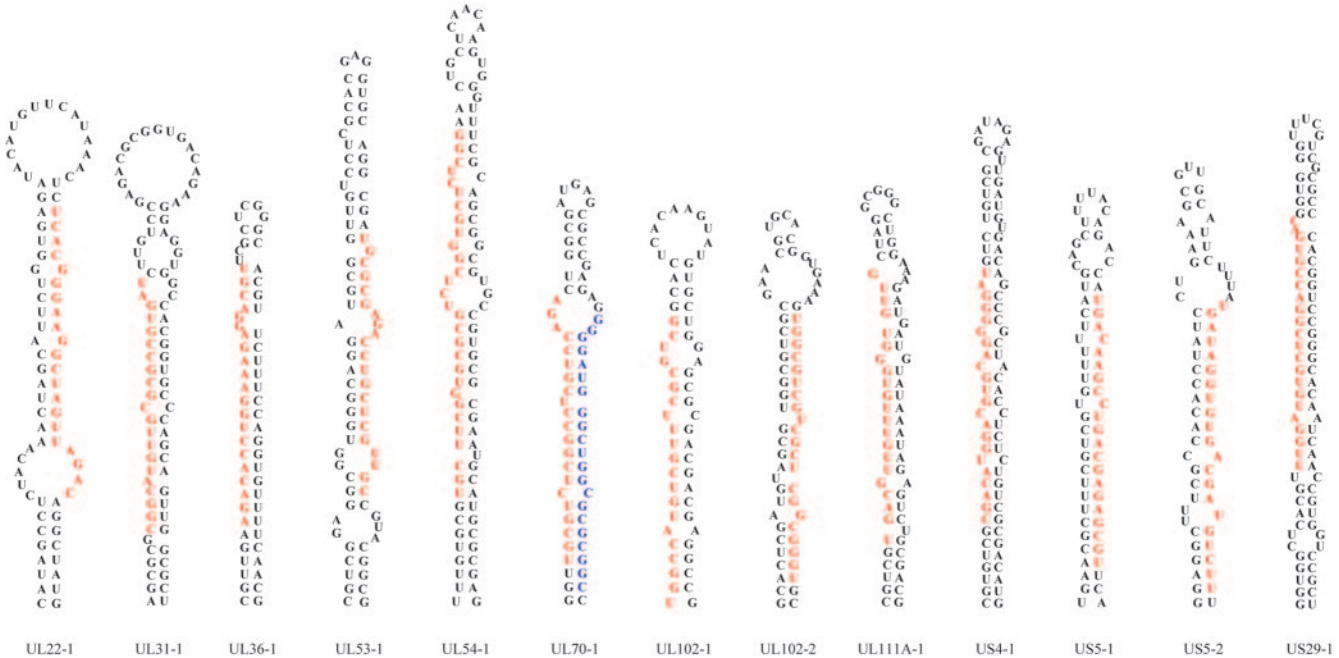

FIG. 1. Predicted stem-loop secondary structures of HCMV-encoded candidate miRNAs. Sequences in red indicate the miRNA sequences predicted by the MiRscan algorithm. For UL70-1, the detected miRNA is indicated in blue. Secondary structures were determined using Mfold (11).

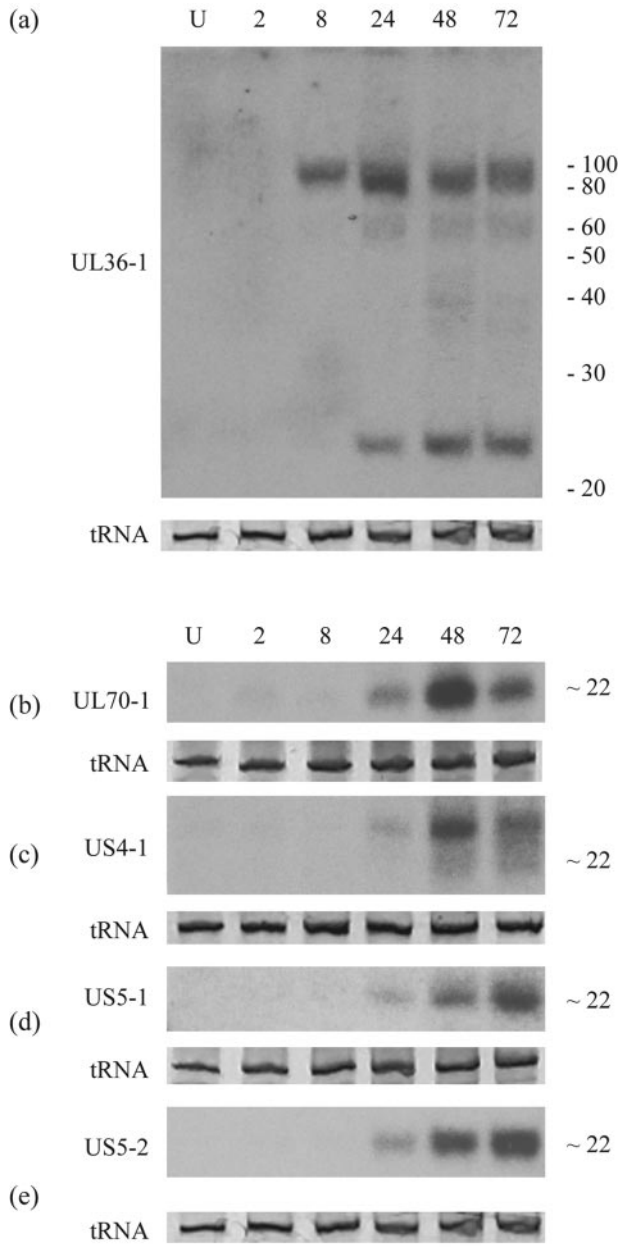

FIG. 2. Northern blot analysis of HCMV-expressed miRNAs. Human fibroblast cells were infected at a multiplicity of 5 PFU per cell. Total RNA was harvested and subjected to Northern blot analysis using probes specific for predicted viral miRNA sequences. (a) Expression of UL36-1 miRNA as well as the ~80-base pre-miRNA species. (b to e) Expression of miRNAs UL70-1, US4-1, US5-1, and US5-2. Ethidium bromide staining of the tRNA band is shown as loading control for each blot. Lane U, uninfected; other lanes are from 2, 8, 24, 48, and 72 h p.i. Positions of radioactive size markers are indicated on the right of the blot.

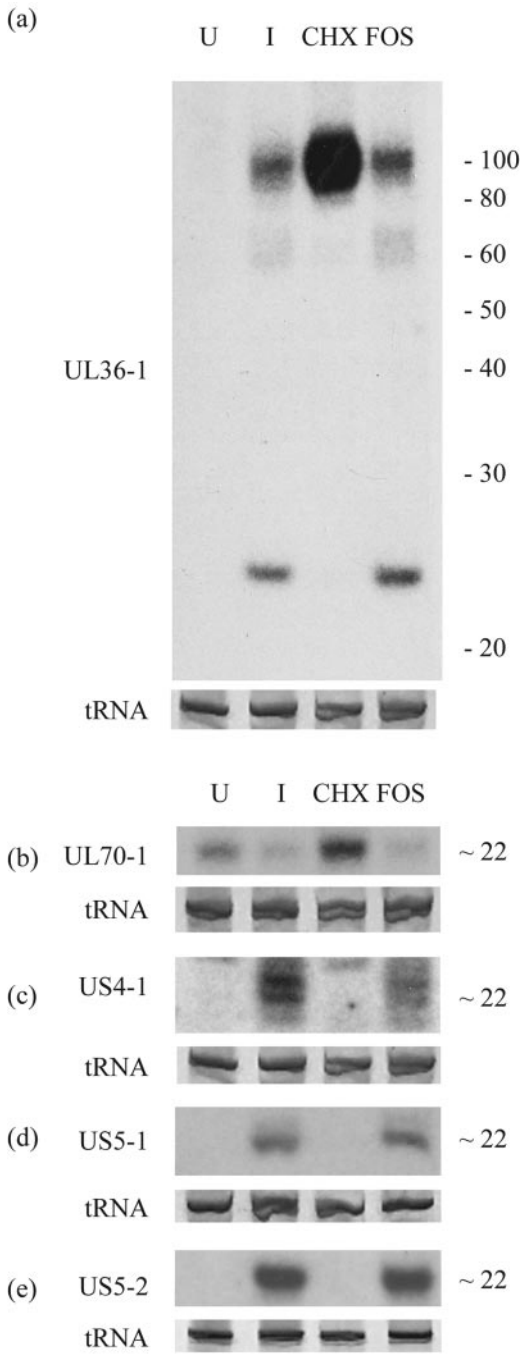

FIG. 3. Expression kinetics of HCMV miRNAs. Human fibroblast cells were treated with either cycloheximide, foscarnet, or with no drug and infected at a multiplicity of 5 PFU per cell. At 36 h, total RNA was harvested and subjected to Northern blot analysis. Total RNA from infected cells was harvested as a negative control. Lanes: U, uninfected; I, infected, no drug; CHX, cycloheximide treated; FOS, foscarnet treated. (a) Expression of UL36-1 miRNA and potential pre-miRNA sequence. (b to e) Expression of UL70-1, US4-1, US5-1, and US5-2. Positions of radioactive size markers are indicated on the right of the blots. tRNA band loading control was visualized by ethidium bromide staining of the polyacrylamide gel.

not detected and was most likely rapidly degraded by cellular enzymes. Interestingly, in the case of UL70-1, the sequence on the complementary strand of the stem-loop corresponding to the predicted miRNA (Fig. 1) was detected rather than the predicted miRNA. This sequence was also first detected at 24 h

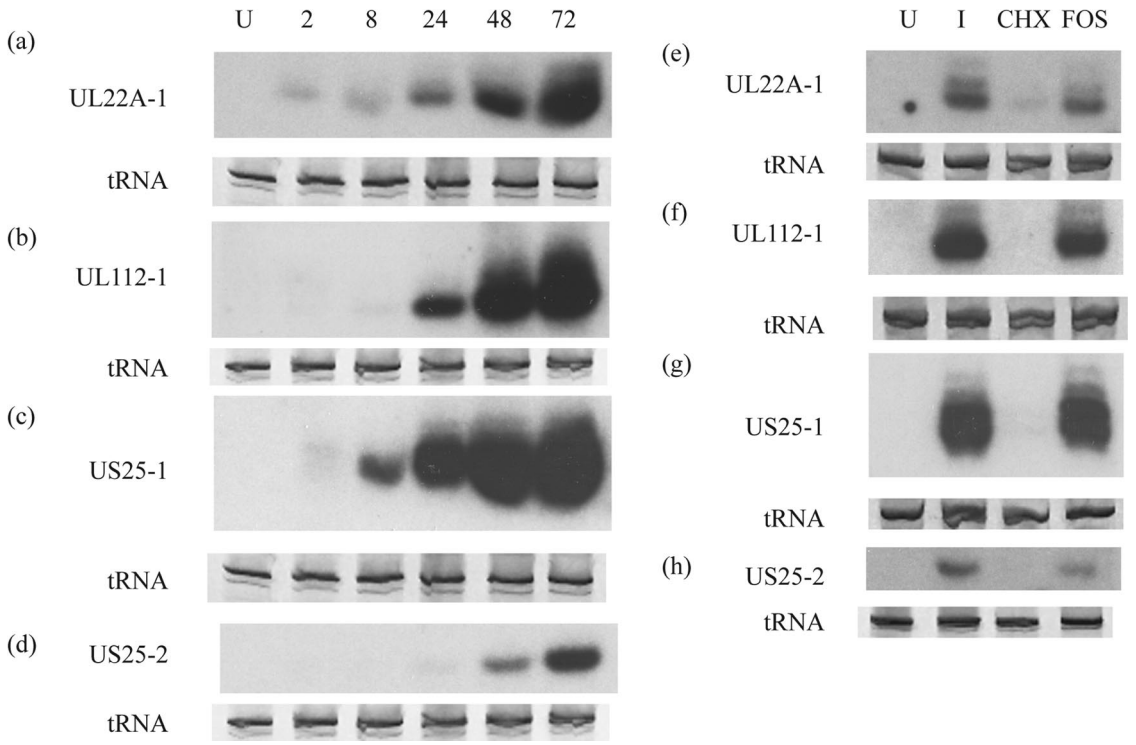

FIG. 4. Northern blot analysis of HCMV-expressed miRNAs. Human fibroblast cells were infected at a multiplicity of 5 PFU per cell. Total RNA was harvested and subjected to Northern blot analysis using probes specific for predicted viral miRNA sequences. (a to d) Expression of miRNAs UL22A-1, UL112-1, US25-1, and US25-2. Ethidium bromide staining of the tRNA band is shown as a loading control for each blot. Lane U, uninfected; other lanes are from 2, 8, 24, 48, and 72 h p.i. (e to h) Expression of miRNAs in the presence of cycloheximide, foscarnet, or no drug. Uninfected controls are shown. Lanes: U, uninfected; I, infected, no drug; CHX, cycloheximide treated; FOS, foscarnet treated.

p.i. but, unlike the other viral miRNAs, peak levels were detected at 48 h p.i. and then dropped to lower levels at 72 h. In each case, signal was not detected using probes to the opposite arms of the stem-loops from which the miRNAs were detected, indicating that the miRNA* species could not be detected by Northern blotting and that the specific bands detected are not due to degraded mRNA transcripts.

Expression of HCMV transcripts can be grouped into three kinetic classes, immediate early, early, or late, based on their requirement for expression of viral protein and viral DNA replication (3). To determine the specific kinetic class of each of the viral miRNAs, viral infections were performed in the presence of either cycloheximide, which blocks protein translation, suggesting early gene expression, or foscarnet, which blocks DNA replication, indicating late gene expression. Total RNA samples were harvested 36 h p.i. followed by Northern blot analysis. Figure 3 demonstrates that cycloheximide but not foscarnet blocked the expression of UL36-1, indicating that UL36-1 is expressed with early kinetics. This observation was surprising, as the UL36 transcript that contains the intron encoding the UL36-1 miRNA exhibits immediate early kinetics (10). Furthermore, the ~80-base species detected in the initial Northern blot assay time course was detected at elevated levels in the cycloheximide sample. One potential explanation for the block in the production of mature miRNAs would be that the extended cycloheximide treatment depleted proteins, such as dicer, required for the processing of mature miRNAs. However, this expression pattern was unique to UL36-1; corresponding species were not detected for UL70-1, US4-1, US5-1, or US5-2 following cycloheximide treatment. Investigation of earlier time points also demonstrated a similar expression pattern for UL36-1 with as little as 14 h of cycloheximide treatment, and levels of the cellular miRNA, miR-22, remained unaffected after 36 h of cycloheximide treatment, suggesting that the cells were still able to produce mature miRNAs (data not shown). Two possible alternative explanations are that the UL36-1 pre-miRNA is transcribed independently rather than being processed from the UL36 intron, or that an early viral gene product is required for the processing of UL36-1 from the pre-miRNA transcript. Expression of US4-1, US5-1, and US5-2 were blocked by cycloheximide treatment but not foscarnet treatment, indicating that these transcripts were also expressed with early kinetics. In contrast, UL70-1 was expressed in the presence of cycloheximide, indicating that this miRNA does not require de novo viral protein expression and is therefore expressed with immediate-early kinetics. The probe for UL70-1 also consistently cross-hybridized to a species in the uninfected control sample. This cross hybridization is unique to the UL70-1 sequence and does not occur with any of the other viral probes. This observation may suggest that UL70-1 has some homology to a cellular miRNA or that the probe is nonspecifically binding to a cellular transcript.

A recent study by Pfeffer et al. (7) used a cloning technique to identify miRNAs encoded by HCMV during acute infection of fibroblast cells. A total of nine miRNAs were identified by cloning, including three miRNAs from the same stem-loops

identified in this study (UL36-1, US5-1, and US5-2). To extend the initial identification of these additional miRNAs, Northern blot analysis was used to determine the expression kinetics during HCMV acute infection. Figure 4 indicates that the virally encoded miRNAs all follow a similar expression pattern, with levels continuing to increase over time. The kinetic studies also indicate that all the miRNAs with the exception of UL70-1 are expressed with early kinetics. We were unable to detect US33-1 by Northern blot analysis.

The computer algorithm recently described by Pfeffer et al. (7) identified sequences with the potential to form stem-loop structures similar to previously identified pre-miRNAs. The sequences were scored and then ranked, and a cutoff score was determined, which reduced false positives. This approach predicted 11 pre-miRNAs, of which 5 were validated by cloning. The approach used in this study not only utilized the characteristic structural features of pre-miRNAs, but also made use of comparative sequence conservation between HCMV and CCMV. This approach was equally effective, with the identification of five validated miRNAs, including two miRNAs, UL70-1 and US4-1, not identified in the Pfeffer et al. study. Of the remaining miRNAs identified by Pfeffer et al. using cloning, all but one of the sequences was conserved in CCMV and scored above or close to 10 using the MiRscan algorithm. Therefore, by refining the SLF algorithm, the success rate of the approach used in this study could be significantly increased in future studies. One possible concern with both bioinformatics approaches is the inability to precisely predict the position of the mature miRNA within the stem-loop structure. Differences in the 5′ and 3′ ends of the sequences identified by cloning compared to the predicted sequences of the miRNAs identified by the MiRscan algorithm suggest that further refinement of the bioinformatics method is required to allow accurate prediction of the exact sequence of the mature miRNA. In conclusion, the data presented show that the approach used is a viable method for the prediction of miRNAs expressed by viral genomes and will be useful in further investigations of viral miRNA identification.

This work was supported by a Public Health Service grant from the National Institutes of Health (AI21640 to J.A.N.) and by a National Science Foundation grant (MCB-0209836 to J.C.C.).

We also acknowledge Alec Hirsch, Michael Jarvis, Dan Streblow, and Cary Rue for useful discussions and Derek Drummond, Patsy Smith, and Heather Meyers for technical assistance.

REFERENCES

1. **Bennasser, Y., S. Y. Le, M. Benkirane, and K. T. Jeang.** 2005. Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. Immunity **22:**607–619.
2. **Cai, X., S. Lu, Z. Zhang, C. M. Gonzalez, B. Damania, and B. R. Cullen.** 2005. Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. Proc. Natl. Acad. Sci. USA **102:**5570–5575.
3. **DeMarchi, J. M., C. A. Schmidt, and A. S. Kaplan.** 1980. Patterns of transcription of human cytomegalovirus in permissively infected cells. J. Virol. **35:**277–286.
4. **Lecellier, C. H., P. Dunoyer, K. Arar, J. Lehmann-Che, S. Eyquem, C. Himber, A. Saib, and O. Voinnet.** 2005. A cellular microRNA mediates antiviral defense in human cells. Science **308:**557–560.
5. **Lim, L. P., N. C. Lau, E. G. Weinstein, A. Abdelhakim, S. Yekta, M. W. Rhoades, C. B. Burge, and D. P. Bartel.** 2003. The microRNAs of Caenorhabditis elegans. Genes Dev. **17:**991–1008.
6. **Pasquinelli, A. E., B. J. Reinhart, F. Slack, M. Q. Martindale, M. I. Kuroda, B. Maller, D. C. Hayward, E. E. Ball, B. Degnan, P. Muller, J. Spring, A. Srinivasan, M. Fishman, J. Finnerty, J. Corbo, M. Levine, P. Leahy, E. Davidson, and G. Ruvkun.** 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature **408:**86–89.
7. **Pfeffer, S., A. Sewer, M. Lagos-Quintana, R. Sheridan, C. Sander, F. A. Grasser, L. F. van Dyk, C. K. Ho, S. Shuman, M. Chien, J. J. Russo, J. Ju, G. Randall, B. D. Lindenbach, C. M. Rice, V. Simon, D. D. Ho, M. Zavolan, and T. Tuschl.** 2005. Identification of microRNAs of the herpesvirus family. Nat. Methods **2:**269–276.
8. **Pfeffer, S., M. Zavolan, F. A. Grasser, M. Chien, J. J. Russo, J. Ju, B. John, A. J. Enright, D. Marks, C. Sander, and T. Tuschl.** 2004. Identification of virus-encoded microRNAs. Science **304:**734–736.
9. **Sullivan, C. S., A. T. Grundhoff, S. Tevethia, J. M. Pipas, and D. Ganem.** 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. Nature **435:**682–686.
10. **Wathen, M. W., and M. F. Stinski.** 1982. Temporal patterns of human cytomegalovirus transcription: mapping the viral RNAs synthesized at immediate early, early, and late times after infection. J. Virol. **41:**462–477.
11. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:**3406–3415.