Commentary

The making of infectious viral RNA: No size limit in sight

Michael M. C. Lai*

Howard Hughes Medical Institute and Department of Molecular Microbiology and Immunology, University of Southern California School of Medicine, 2011 Zonal Avenue, Los Angeles, CA 90033-1054

Despite the disadvantage of having an RNA genome, which is more difficult than DNA to be genetically tinkered, the reverse genetics of RNA viruses actually originated at about the same time as the dawning of the genomic manipulation of DNA viruses. The first RNA virus to be genetically modified was Qβ phage (1). Initially, the RNA molecules were chemically modified during RNA replication *in vitro*; the procedures were cumbersome and the range of RNA mutations was limited. Nevertheless, the potential power of reverse genetics as a tool for studying RNA viruses was transparently clear in a pioneering series of site-specific mutagenesis studies from C. Weissmann's laboratory (2, 3). The advent of recombinant DNA technology in the 1970s prompted RNA virologists to convert viral RNA genomes into complementary DNA copies and replicate them as plasmid inserts in bacterial hosts for easier genetic manipulation. Amazingly, the plasmid containing the complete cDNA of the Qβ phage RNA was fully infectious when introduced into bacterial hosts and was capable of completing the full viral replication cycle (4). Presumably, transcription of the viral RNA was randomly initiated, and the RNA was processed mysteriously to the correct viral sequence. Later, this technique was applied to several other viruses, including poliovirus (5) and viroids (6). Infectious poliovirus cDNA constructs remained the staples of poliovirus genetics for many years after that. Subsequently, another technique was developed whereby RNA was made by *in vitro* transcription of viral cDNA templates linked to a promoter recognized by *Escherichia coli* or phage DNA-dependent RNA polymerases (7). When the RNA transcribed *in vitro* was transfected into cells, it led to viral RNA replication (8). The first virus thus studied was brome mosaic virus, a relatively small plant virus containing three RNA segments of 3.2, 2.8, and 2.1 kb. In contrast to the cDNA transfection approach, the RNA molecules generated by using the RNA transfection approach were engineered so that they had well defined ends that matched the natural viral RNA sequences. This elegant approach empowered the virologists working with viruses containing relatively small RNA genomes or multiple RNA segments with the tools of reverse genetics. In all of these approaches, the size of the viral RNA was a major limitation. The Qβ phage RNA is 4.5 kb, and the poliovirus RNA is 7.5 kb. Over time, these approaches have been refined to enable the cloning of progressively larger RNAs. With some exceptions, most viral RNAs up to 15 kb long can now realistically be cloned. The report by Almazán *et al.* in this issue of PNAS (9) represents a further quantum leap, i.e., the successful cloning of a 27-kb long RNA derived from a coronavirus porcine transmissible gastroenteritis virus (TGEV), a task previously thought to be unachievable. This accomplishment is an intellectual and engineering tour de force. Because coronavirus contains the longest viral RNA genome by far (and is probably one of the longest stable RNAs in nature), this approach seems to pave the way for the reverse genetics studies for all RNA viruses.

Long Viral cDNAs: Problems and Solutions. The large size of some viral RNAs presented several obstacles to constructing an infectious cDNA or RNA transcript. First, long RNA sequences make the synthesis of a faithful cDNA molecule difficult, because the fidelity of reverse transcriptase–PCR (RT-PCR) for the amplification of cDNA inevitably decreases in proportion to the RNA length. This difficulty is compounded by the quasispecies nature of RNA viruses (i.e., viral RNA consists of multiple RNA sequences with minor sequence variations). Second, long RNA sequences are more likely to contain fortuitous poison sequences, which make the cDNA sequence in plasmids unstable. Third, it is difficult to find a suitable vector that can accommodate large foreign cDNA inserts.

The first difficulty has been overcome largely by the improvement of RT-PCR procedures. The availability of high-fidelity RT and polymerases has significantly decreased the error rate of RT-PCR. Even so, laborious procedures usually are required to correct the cDNA sequences so that they match the consensus sequence of the viral RNA, because the PCR products often reflect minor and defective RNA sequences present in the virus population. The recent success in the cloning of infectious hepatitis C virus RNAs (10, 11) best illustrates the necessity of this step. The second difficulty, the presence of poison sequences in the cDNA, is a particularly irksome problem in the cloning of viral cDNA or DNA, probably because bacteria have not been adapted to such foreign sequences. Bacteria also have the capacity to artificially select particular viral sequences; thus, the cloned sequences obtained often are nonrandomly biased rather than representative of the majority RNA sequences (12). Solutions to the poison sequence problems have been made previously. For example, the cDNA copy of yellow fever virus RNA could not be cloned in one piece; therefore, it was cloned in two segments and then ligated *in vitro* to make a full-length cDNA for *in vitro* transcription (13). Thus, the passage of poison sequences in bacteria was avoided. This approach has been adapted for the rapid cloning of flavivirus RNA in general (14). The third difficulty has been partially overcome by the use of various vector systems. Bacterial artificial chromosome (BAC), which was used successfully for cloning the 150-kbp herpes simplex virus DNA (15), is touted in this report (9) as another versatile cloning vector.

Travails and Triumph of Cloning an Infectious Coronavirus Genome. Variations of the cloning methods have made possible the construction of infectious cDNA or RNA for most RNA viruses. However, for obvious reasons, one remaining virus that has so far resisted the onslaught of cloning attempts is coronavirus. Coronaviruses include many economically and medically

See companion article on page 5516.

*E-mail: michlai@hsc.usc.edu.

important viruses, e.g., porcine TGEV, mouse hepatitis virus (MHV), avian infectious bronchitis virus (IBV), and human coronaviruses, the last of these being responsible for many common colds and, possibly, gastroenteritis and neurological illnesses, such as multiple sclerosis (16). These viruses contain a positive-sense RNA genome of 27–32 kb, which is more than twice the size of the largest genomic RNA of the conventional RNA viruses. The viral RNA is replicated by an RNA-dependent RNA polymerase entirely in the cytoplasm, independently of the nucleus, although recent studies have suggested that nuclear factors are involved in viral RNA synthesis (17). More than two-thirds of the viral RNA sequences are devoted to making gene products involved in viral RNA synthesis. The enormous size of the coronaviral RNA defies the theoretical predictions of the upper limit of RNA size, based on the high error frequencies and the lack of proofreading activities of RNA polymerases (18). The fact that the 27- to 32-kb long coronavirus RNA can be stably maintained is thought to be in part attributable to the high frequency of RNA recombination (19). Whatever the reason, the large size of the coronavirus RNA posed a daunting obstacle to the construction of an infectious cDNA or RNA for these viruses. Almazán and colleagues (9) have solved these problems with two significant innovations, suggesting not only a versatile approach to cloning long viral cDNAs, but also possible new ways of overcoming the host's restriction on foreign RNA molecules.

The first step described in Almazán and colleagues' report is overcoming the effects of poison sequences. The authors achieved this by cloning the region containing the poison sequences in the last cloning step before the whole sequence was inserted into a BAC, which apparently can tolerate more exogenous sequences than other vectors. This simple trick precluded the possible deleterious effects of the poison sequences inevitably present in long cDNA sequences.

The second innovative, and somewhat surprising, feature of this study is the resurrection of the old trick of using cDNA transfection to drive the production of viral RNA *in situ* (Table 1), but with one notable difference. In previous cDNA transfection studies, the viral cDNA was expressed by random initiation of transcription, probably from within the plasmid sequences; however, in the studies reported here, the viral cDNA was placed under a specific promoter (cytomegalovirus immediate-early promoter), and the ends of viral RNA were carefully engineered to match their natural sequences. In any case, the cDNA transfection approach was intuitively thought to be artificial because the viral RNAs in question (e.g., poliovirus and coronavirus) naturally replicate only in the cytoplasm; thus, the production of viral RNA in the nucleus would introduce additional roadblocks to the replication of viral RNA in the cytoplasm. Past examples of the successful use of the cDNA transfection approach for RNA viruses mostly involved viruses that normally replicate in the nucleus, such as hepatitis delta virus (20), viroids (6), and, more recently, influenza virus (21). For cytoplasmic viruses (e.g., most of the negative-strand RNA viruses), an alternative cDNA transfection approach was developed in which viral cDNA (under a T7 promoter) transfection is coupled with the expression of T7 polymerase through a recombinant vaccinia virus, which replicates in the cytoplasm (refs. 22 and 23; Table 1). Such an approach ensures that viral RNA is transcribed directly in the cytoplasm, where it replicates. Nonetheless, the original success with poliovirus cDNA (5) and the recent successes with influenza virus cDNA (under a polymerase I promoter) and, now, coronavirus cDNA (under a pol II-mediated promoter) will likely challenge our stereotypic reservations regarding the appropriateness of the atopic expression of viral RNA in a different subcellular compartment.

Table 1. Strategies for making infectious viral RNA or cDNA

Transfectants	Sites of primary RNA transcripts	Successful examples
cDNA in plasmids (no promoters)	Nucleus	Poliovirus (C), viroids (N)
cDNA under pol I or pol II promoters	Nucleus	Influenza virus (N), coronavirus (C)
In vitro RNA transcripts	Cytoplasm	Numerous positive-stranded RNA viruses (C)
cDNA under phage (T7) promoter + vaccinia virus-T7 polymerase	Cytoplasm	Numerous negative-stranded RNA viruses (C)

N, nuclear replication; C, cytoplasmic replication.

The Surprises of Atopic RNA Expression in the Nucleus. A very surprising finding of this study is that the coronaviral RNA, which is normally present only in the cytoplasm, is not spliced in the nucleus and is successfully exported as an intact molecule to the cytoplasm. The occurrence of viral RNA splicing in this case would have been the antithesis to the synthesis of infectious viral RNA. Almazán and colleagues (9) have shown that the 27-kb TGEV RNA contains multiple consensus-splicing signals, and yet it was not significantly spliced in the nucleus. Whether this finding is the luck of the draw or a general phenomenon for foreign RNAs remains to be seen. In any case, it raises the hope that perhaps RNA splicing may not be the stumbling block to the expression of foreign sequences in the nucleus as previously feared. Even more surprising is that the viral RNA is successfully exported to the cytoplasm despite the absence of splicing. This finding brings to mind a recent study showing that RNA splicing and the export of RNA from the nucleus are coupled (24). However, the success of the DNA transfection approach for making infectious coronaviral RNA suggests that such a coupling may not apply to all foreign RNA sequences. Perhaps other RNA motifs govern RNA export.

The efficiency of the generation of infectious virus particles by this approach was understandably low. This study did not compare the efficiencies of conventional RNA transfection with the DNA transfection methods. Is DNA transfection really necessary? How does it compare with RNA expression directly in the cytoplasm? RNA expression from the DNA template in the nucleus may yield a large amount of RNA; however, it will be counterbalanced by the quality control mechanisms (i.e., degradation) imposed on the nonprocessed foreign RNAs. In contrast, direct RNA expression in the cytoplasm (by RNA transfection or transcription by T7 polymerase), although inefficient, may cause RNA to accumulate in the cytoplasm, where it replicates (Table 1). Further studies will be required to determine the relative merits of these approaches. In any case, the reported success of cloning and expressing such a large cDNA fragment surely will inspire others to further improve the methodology, perhaps by developing an efficient cytoplasmic expression vehicle with a large cloning capacity.

Reverse Genetics for All (Almost). The successful cloning of an infectious TGEV cDNA is, of course, an important breakthrough for coronavirus research. The Almazán and colleagues' (9) study showcases the potential power of reverse genetics that has eluded coronavirus researchers until now. The data presented here clearly show that the spike protein alone is sufficient to determine the pathogenicity of the virus, thus explaining the

mechanism by which a porcine respiratory coronavirus emerged from the enterotropic TGEV in Europe and the U.S. in the early 1980s (25, 26). Further insights into the molecular basis of viral pathogenesis will now be possible with the availability of the infectious cDNA.

Except for a few fellow coronaviruses that are yet to be conquered (e.g., MHV with a 32-kb genome), the vista for the cloning of long viral RNA appears to be clear, now that the TGEV RNA has been successfully cloned. This study (9), together with the previous reports of the successful cloning of negative-stranded RNA viruses (22) and segmented RNA viruses (21), essentially establishes the experimental paradigms for the cloning of the genomes of all classes of single-stranded RNA viruses. Double-stranded RNA viruses may not be that far behind. It is now high time to exploit the reverse genetics of all RNA viruses, an undertaking once thought impossible. This same cloning approach also may be applicable to the expression of long cellular RNAs.

1. Flavell, R. A., Sabo, D. L., Bandle, E. F. & Weissmann, C. (1974) *J. Mol. Biol.* **89,** 255–272.
2. Domingo, E., Flavell, R. A. & Weissmann, C. (1976) *Gene* **1,** 3–25.
3. Sabo, D. L., Domingo, E., Bandle, E. F., Flavell, R. A. & Weissmann, C. (1977) *J. Mol. Biol.* **112,** 235–252.
4. Taniguchi, T., Palmieri, M. & Weissmann, C. (1978) *Nature (London)* **274,** 223–228.
5. Racaniello, V. R. & Baltimore, D. (1981) *Science* **214,** 916–919.
6. Cress, D. E., Kiefer, M. C. & Owens, R. A. (1983) *Nucleic Acids Res.* **11,** 6821–6835.
7. Ahlquist, P. & Janda, M. (1984) *Mol. Cell. Biol.* **4,** 2876–2882.
8. Ahlquist, P., French, R. & Bujarski, J. J. (1987) *Adv. Virus Res.* **32,** 215–242.
9. Almazán, F., González, J. M., Pénzes, Z., Izeta, A., Calvo, E., Plana-Durán, J. & Enjuanes, L. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5516–5521.
10. Kolykhalov, A. A., Agapov, E. V., Blight, K. J., Mihalik, K., Feinstone, S. M. & Rice, C. M. (1997) *Science* **277,** 570–574.
11. Yanagi, M., Purcell, R. H., Emerson, S. U. & Bukh, J. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 8738–8743.
12. Forns, X., Bukh, J., Purcell, R. H. & Emerson, S. U. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 13909–13914.
13. Rice, C. M., Grakoui, A., Galler, R. & Chambers, T. J. (1989) *New Biol.* **1,** 285–296.
14. Gritsun, T. S. & Gould, E. A. (1995) *Virology* **214,** 611–618.
15. Messerle, M., Crnkovic, I., Hammerschmidt, W., Ziegler, H. & Koszinowski, U. H. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 14759–14763.
16. Lai, M. M. C. & Cavanagh, D. (1997) in *Advances in Virus Research* (Academic, New York), pp. 1–100.
17. Lai, M. M. C. (1998) *Virology* **244,** 1–12.
18. Reanney, D. C. (1982) *Annu. Rev. Microbiol.* **36,** 47–73.
19. Lai, M. M. C. (1992) *Microbiol. Rev.* **56,** 61–79.
20. Kuo, M. Y.-P., Chao, M. & Taylor, J. (1989) *J. Virol.* **63,** 1945–1950.
21. Neumann, G., Watanabe, T., Ito, H., Watanabe, S., Goto, H., Gao, P., Hughes, M., Perez, D. R., Donis, R., Hoffmann, E., *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9345–9350.
22. Schnell, M. J., Mebatsion, T. & Conzelmann, K.-K. (1994) *EMBO J.* **13,** 4195–4203.
23. Roberts, A. & Rose, J. K. (1998) *Virology* **247,** 1–6.
24. Luo, M.-J. & Reed, R. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 14937–14942.
25. Pensaert, M., Callebaut, P. & Vengote, J. (1986) *Vet. Q.* **8,** 257–261.
26. Wesley, R. D., Woods, R. D., Hill, H. T. & Biwer, J. D. (1990) *J. Vet. Diagn. Invest.* **2,** 312–317.