The virology–RNA biology connection

BRYAN R. CULLEN

Department of Molecular Genetics and Microbiology and Center for Virology, Duke University Medical Center, Durham, North Carolina 27710, USA

My goal in writing this article is to offer my personal reflections on RNA research, focusing on the last 20 years, to be published in a special issue marking the 20th anniversary of *RNA*, the premier journal in the then-emerging but now well established field of RNA biology. While I have published a number of papers relevant to RNA research, my main focus has always been on virology and I will therefore take this opportunity to remind especially younger readers that RNA research actually initially grew out of virology and that these two subjects continue to be highly interconnected.

Viruses contain small genomes, which can be more readily manipulated than cellular genomes and, in infected cells, viruses often produce very high levels of individual viral transcripts, which can be more readily isolated and characterized than most cellular RNAs. It is therefore not surprising that many of the key post-transcriptional modifications that affect mammalian mRNAs were first defined using viral systems. The capping of eukaryotic mRNA was first demonstrated in reovirus and vaccinia virus in 1975 and the role of capping in promoting mRNA translation first demonstrated that same year using viral mRNAs. Splicing, and then alternative splicing, were first demonstrated by analysis of late adenovirus transcripts and study of these same, highly expressed mRNAs also revealed the first example of alternative poly(A) site usage and led to the demonstration that poly(A) addition sites are generated by endonucleolytic cleavage, rather than by termination. Analysis of viral systems, in particular picornaviruses, led to the first description of an Internal Ribosome Entry Site and analysis of retroviral replication also led to the discovery of regulated translational frameshifting.

Retroviruses also proved to be a key experimental system to study nuclear mRNA export due to their unique requirement for the export of both spliced and unspliced forms of the same genome-length transcript. Cells normally block the nuclear export of cellular pre-mRNAs that retain intact splice sites and incompletely spliced retroviral RNAs are therefore unable to access the canonical nuclear mRNA export pathway and must instead be exported via a distinct mechanism. The first nuclear mRNA export factor to be defined, HIV-1 Rev, recruits a cellular nuclear export factor called Crm1 to a stem–loop structure, the Rev Response Element (RRE), present on incompletely spliced HIV-1 transcripts. The binding site for Crm1 on Rev, a leucine-rich motif that is essential for Rev function, was the first nuclear export signal (NES) to be defined and remains the prototype of the most common class of NESs.

While Crm1 can be recruited by Rev to the RRE to induce HIV-1 mRNA nuclear export, Crm1 does not normally function as a generic mRNA export factor and instead mediates the nuclear export of non-coding RNA ribonucleoprotein complexes and proteins. The primary cellular mRNA nuclear export factor, NXF1, was first identified as the protein that mediates the nuclear export of incompletely spliced RNAs expressed by a different retrovirus, Mason Pfizer Monkey Virus (MPMV). MPMV does not encode a Rev homolog and instead contains an RNA stem–loop structure, the constitutive transport element (CTE), that is necessary and sufficient for nuclear export of unspliced MPMV transcripts and that can functionally substitute for the HIV-1 Rev/RRE axis. The CTE directly interacts with NXF1 and this short-circuits the normally tightly regulated recruitment of NXF1 to fully spliced cellular mRNAs to allow the nuclear export of unspliced MPMV RNAs.

The 20 years prior to 1995 in many ways represented a "golden age" of RNA research in that many of the key steps in the processing and function of mRNAs were defined during this period. In contrast, the subsequent 20 years have in many ways been focused on defining the molecular mechanisms that regulate and/or mediate these processing steps. In particular, there has been considerable progress in understanding the regulation of alternative splicing and polyadenylation, which together are responsible for much of the complexity of eukaryotic gene expression. Nevertheless, there have also been several major discoveries in RNA research in the last 20 years that merit a more detailed discussion.

RNA interference

RNA interference (RNAi) was initially described in nematodes in 1998 as a process whereby exogenous long double-stranded RNAs (dsRNAs) induce the selective degradation of cognate mRNA species. During RNAi, long dsRNAs are cleaved by the processive exonuclease Dicer to yield short, ~22 bp small interfering RNAs (siRNAs). All RNA viruses, except retroviruses, generate long dsRNAs during their replication cycle and most DNA viruses undergo convergent transcription of both strands of their genome, which also generates long dsRNAs. During RNAi, these long viral dsRNAs could be processed by Dicer to yield protective siRNAs that might inhibit virus replication and RNAi indeed clearly functions as a key antiviral innate immune response in plants and invertebrates. It is therefore thought that RNAi initially evolved as an innate antiviral immune response and, in fact, one of the first demonstrations of siRNAs of any kind was in plants infected by potato virus X, where high levels of ~22 nt RNAs of viral origin were observed. In contrast, in mammalian somatic cells, long viral dsRNAs are not subject to Dicer cleavage and virus-derived siRNAs are therefore not made. Instead, these long viral dsRNAs induce the interferon response, a protective innate immune response mediated by an array of cellular proteins. Why RNAi has ceased to function as an antiviral mechanism in mammals, yet remains an important antiviral response in plants and invertebrates, is a fascinating question that remains unresolved.

MicroRNAs

While mammalian Dicer has lost the ability to process long dsRNAs into siRNAs, it remains fully capable of processing pre-miRNA stem–loops into mature microRNAs (miRNAs) and we now know that mammals encode large numbers of miRNAs that play a critical role in many aspects of cellular differentiation and development. MiRNAs, while normally encoded within the host cell genome, are functionally closely analogous to siRNAs in that they are loaded into the same RNA-induced silencing complexes (RISCs) and inhibit the expression of complementary mRNA species via the same mechanisms.

While miRNAs were first identified in nematodes, advances in small RNA sequencing techniques soon revealed that all metazoan species express numerous distinct miRNAs. Moreover, several viruses, including almost all herpesvirus species, encode miRNAs that contribute to aspects of viral replication and pathogenesis. In the case of hepatitis C virus (HCV), a hepatocyte-specific cellular miRNA, miR-122, has emerged as an essential co-factor for HCV replication, and antisense drugs that block miR-122 function also effectively block HCV replication in vivo. These drugs may become the first clinically useful agents that exert a beneficial effect by repressing miRNA function. Nevertheless, most viruses are able to replicate normally in cells that, due to mutational inactivation of Dicer, produce no miRNAs whatsoever, thus revealing that miRNAs not only are not required as co-factors for most viral replication cycles but also that viruses are resistant to inhibition by endogenous miRNAs. How viruses systematically avoid inhibition by endogenous cellular miRNAs remains an intriguing question, given that viral RNAs contain numerous potential, computer-predicted miRNA target sites.

CRISPR/Cas RNA-guided DNA endonucleases

The discovery of RNAi led to the suggestion that siRNAs would soon emerge as a novel class of drugs and it was certainly expected that RNAi-based screens would soon lead to a comprehensive identification of genes that promote or retard a wide range of human diseases. In fact, siRNAs have been somewhat disappointing as drugs, due primarily to delivery issues but also due to concerns about off-target effects and the generally transient nature of RNAi. RNAi-based screens have certainly been very informative, and have led to the identification of a number of factors important for disease progression, but even here problems of reproducibility exist, due perhaps to the variable efficacy of different siRNAs in different experimental contexts.

An alternative and potentially very powerful approach has recently emerged with the discovery and analysis of RNA-guided DNA endonucleases belonging to the bacterial CRISPR/Cas system of bacterial adaptive antiviral immunity. In Type II CRISPR/Cas systems, a single effector protein, called Cas9, is recruited to specific DNA target sites by a fully complementary guide RNA. Once bound, Cas9 induces a double-stranded DNA break that, after error prone repair by non-homologous end joining, can result in the stable inactivation of a target gene, rather than the transient knockdown characteristic of RNAi. Indeed, several studies describing efficient gene knockouts in vivo, or the use of Cas9 and libraries of guide RNAs to allow large scale genetic screens, have recently been reported. In my view, CRISPR/Cas systems represent a transformative technology that has enormous potential and that is likely to revolutionize biomedical research. Moreover, because the Cas9 protein and guide RNAs can be readily introduced into cells using viral vectors, they also have the potential to be used for gene repair and for the in vivo elimination of DNA targets, e.g., DNA virus genomes, that are deleterious. I therefore believe that this technology will become a major factor in driving progress in RNA research over the next decade.

Long intergenic non-coding RNAs

Cells express large numbers of non-coding RNAs that play critical roles in aspects of mRNA translation and splicing and that are generally transcribed by RNA polymerase I (pol I) or pol III. It has been clear for some time that both cells and viruses also express long intergenic non-coding RNAs (lincRNAs), transcribed by pol II, that exert a number

of diverse phenotypes and the recent development of deep sequencing technologies has now revealed the expression of hundreds of additional lincRNAs of currently unknown function. Some lincRNAs are nuclear in localization and could, for example, regulate chromatin function, while other lincRNAs are predominantly cytoplasmic and could function as scaffolds or as regulators of miRNA function. At this stage, the functions of very few lincRNAs are understood and the next decade will undoubtedly reveal numerous as yet unsuspected activities. Unfortunately, it is very difficult to deduce what a lincRNA might do, and how it does it, from first principles. For example, the HSV-1 latency associated transcript LAT, a nuclear lincRNA first identified in 1991, is known to regulate viral latency yet how it does so remains unresolved. However, knockouts of cellular lincRNAs can now be rapidly generated using CRISPR/Cas, so that informative phenotypic clues can hopefully be generated. Certainly, a more comprehensive understanding of lincRNA function will require new insights that currently remain elusive.

Conclusions

While virology gave rise to many of the early discoveries in RNA research, this initially close relationship has perhaps become more distant over time. Yet, I would argue that virology provides experimental systems that still have much value in RNA biology, and key RNA-based processes discovered in the recent past, including RNAi and CRISPR/Cas, initially evolved as antiviral immune responses in eukaryotes and bacteria, respectively. I therefore anticipate that virology and RNA biology will each continue to inform progress in the other area of research and will continue to benefit from this cross-fertilization.

Acknowledgments