CHAPTER 4
Preparation and Analysis of RNA

INTRODUCTION

The ability to isolate clean intact RNA from cells is essential for experiments that measure transcript levels, for cloning of intact cDNAs, and for functional analysis of RNA metabolism. RNA isolation procedures frequently must be performed on numerous different cell samples, and therefore are designed to allow processing of multiple samples simultaneously. This chapter begins by describing several methods commonly used to isolate RNA, and concludes with methods used to analyze RNA expression levels, RNA synthesis rates, and genome-wide location of RNAs.

The difficulty in RNA isolation is that most ribonucleases are very stable and active enzymes that require no cofactors to function. The first step in all RNA isolation protocols therefore involves lysing the cell in a chemical environment that results in denaturation of ribonuclease. The RNA is then fractionated from the other cellular macromolecules under conditions that limit or eliminate any residual RNase activity. The cell type from which RNA is to be isolated and the eventual use of that RNA will determine which procedure is appropriate. No matter which procedure is used, it is important that the worker use care (e.g., wearing gloves) not to introduce any contamination that might include RNase during work-up of the samples, and particularly when the samples are prepared for storage at the final step.

While the RNA isolation protocols describe methods that can be performed using common laboratory reagents, several kits for RNA isolation are commercially available. These kits offer the dual advantage of ease of use and (at least in theory) of reagents that have been tested for effectiveness. These kits frequently work well and are widely used. The disadvantages of using kits are that they are more expensive per sample than isolations that are done using "home made" solutions, and that the kits do not offer flexibility for cell types that require special conditions. The cost disadvantage is frequently outweighed in situations where only a few RNA isolations are performed; however, preparing reagents from scratch can take time, and in the event that any of the reagents are not working properly, troubleshooting will require further time. In situations where numerous samples are routinely processed, significant cost savings can be realized by avoiding the use of kits.

One of the primary uses of RNA isolation procedures is the analysis of gene expression. In order to elucidate the regulatory properties of a gene, it is necessary to know the structure and amount of the RNA produced from that gene. The second part of this chapter is devoted to techniques that are used to analyze RNA. Procedures such as S1 nuclease analysis and ribonuclease protection can be used to do fine-structure mapping of any RNA. These techniques allow characterization of 5′ and 3′ splice junctions as well as the 5′ and 3′ ends of RNA. Both of these procedures, as well as northern analysis, can also be used to accurately determine the steady-state level of any particular message.

After determining the steady-state level of a message, many investigators wish to examine whether that level is set by the rate of transcription of the gene. Alterations in steady-state level might also reflect changes in processing or stability of the RNA. *UNIT 4.10* describes the "nuclear run-off" technique, which determines the number of active RNA polymerase

Current Protocols in Molecular Biology 4.0.1-4.0.2, April 2012
Published online April 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/0471142727.mb0400s98

molecules that are traversing any particular segment of DNA. This procedure is used to analyze directly how the rate of transcription of a gene varies when the growth state of a cell is changed.

At times, investigators characterizing active RNA polymerase will desire resolution finer than the 100 bp resolution that can be obtained with nuclear run-on technology. *UNIT 4.14* presents a technology, NET-seq, that is able to map transcribing RNA polymerase with base-pair resolution. This protocol allows localization of transcribing RNA polymerase throughout the yeast genome and can be used to investigate how regulatory changes or mutations impact the location of transcribing RNA polymerase. This can provide insight into regulation of gene sets and into events that occur during elongation of transcripts by RNA polymerase.

Advances in sequencing technologies have resulted in techniques that allow an investigator to sample millions of individual RNAs in a preparation of cells and to use that information to describe the 'transcriptome' of that cell population (*UNIT 4.11*). This technology allows mapping of start sites and of splice sites of RNAs, and also provides quantitative information concerning relative expression levels of RNAs. The ability to sample the entire transcriptome of a given cellular state not only allows the discovery of potentially important novel RNAs but also allows expression-level changes to be examined in a non-biased manner.

A second use of the advances in high throughput sequencing technology is to identify previously undescribed RNA species in the cell. There are numerous non-coding RNAs made in the cell, and many of these produce RNA species, such as microRNAs, that play important regulatory functions. To discover these RNAs, which might be expressed at low levels, methods that remove highly expressed mRNA species are desirable as the large number of those species dominates any high throughput sequencing reaction done on a quantifiably representative library of RNA species. A protocol to digest abundant RNA species prior to sequencing uses Crab duplex-specific nuclease and is described in *UNIT 4.12*. This protocol exchanges the ability to perform quantifiable RNA sequencing studies, as described in *UNIT 4.11*, for the ability to detect a large number of different RNA species made in the cell. It provides an important discovery tool for novel RNA species, both non-coding and mRNAs that are expressed at very low levels.

Frequently, investigators wish to focus on a subset of RNA species involved in a regulated process. In these instances, sequencing the entire transcriptome is not only unnecessary, but complicates and weakens analysis. Thus, methodologies that allow rapid analysis of a specific set of genes are powerful as they allow more facile characterization of how the level of expression of the gene set changes in different cells under different growth conditions and/or developmental stages. *UNIT 4.13* describes a protocol, termed RASL-seq, that allows quantification of a subset of genes in a manner that can be performed on multiple samples and can potentially be automated. This technology trades the ability to obtain an unbiased look at the entire genome for the ability to look at a significant number of genes under a wide variety of settings.

Robert E. Kingston