CHAPTER 23

PROKARYOTIC REVERSE TRANSCRIPTASES

BERT C. LAMPSON*
Department of Health Sciences, East Tennessee State University, Johnson City, TN, USA
*lampson@mail.etsu.edu

1. INTRODUCTION

The reverse transcription of the genetic code of an RNA molecule back into a complementary DNA copy (cDNA) is usually accomplished by a specially dedicated RNA-dependent DNA polymerase called a reverse transcriptase (RT). Reverse transcription is presumed to have been a central event in the transition from the theorized ancient “RNA world” of life, in which RNA molecules served as both the source of genetic information, as well as, catalytic functions for cellular life, to the present day DNA-RNA-protein world. Even today reverse transcription continues to occur in most organisms from human cells to bacteria. For example, a wide assortment of genetic elements found in plant, animal, and microbial cells use reverse transcription for at least part of their replication or mobility. These include RNA viruses, DNA viruses, transposons, introns, and mitochondrial plasmids (Eickbush 1994; Eickbush and Malik 2002). More astonishing than the wide variety of retroelements that encode a RT is the colossal number of repetitive retrosequences, such as the Alu sequences, that have been generated by these elements in many eukaryotic genomes. These retrosequences are DNAs that usually do not code for RT but have clearly been produced by reverse transcription of an RNA molecule (Kazazian 2004). In addition to the production of these “parasitic” DNAs, reverse transcription also serves a vital function for most eukaryotic organisms. Here the essential enzyme telomerase, which functions to maintain the telomere ends of chromosomes, is a type of RT (Lue 2004).

Most RTs are placed in two very broad categories based on the phylogenetic analysis of their amino acid sequence and the type of retroelement that codes for these polymerases. The first group is found in eukaryotic cells and is usually called LTR-containing retroelements because their DNA is flanked by long terminal repeat.
sequences. These elements include retroviruses and the virus-like retrotransposons called Ty in yeast cells. The second group is called non-LTR retroelements because they are not flanked by long terminal repeats and are a very diverse collection of elements found in both eukaryotic and prokaryotic organisms. These include the mobile group II introns, the L1 retrotransposons, and the various retroelements found in bacteria. Telomerase is phylogenetically related to the RTs from this group and is thus also included in this second category (Eickbush and Malik, 2002).

Among the best characterized RTs are the polymerases from retroviruses and include the detailed crystal structure of the RT from the HIV-1 virus (Steitz, 1999). The focus of this chapter is the recently discovered RTs found in bacteria and the interesting genetic elements that encode them. These prokaryotic RTs are considered to be the ancestors of all retroelements found in eukaryotic genomes based on phylogenetic comparisons (Toor et al., 2001). In addition, the bacterial RTs are a very diverse group of proteins with a number of novel properties. Thus, these bacterial RTs represent an emerging new resource for many potential nucleic acid based technologies like RT-PCR.

Prokaryotic RTs generally fall into one of three different types depending on the type of retroelement DNA that codes for these proteins. These groups or types are also in agreement with phylogenetic groupings for RTs determined by comparing their amino acid sequences. The three types of retroelements are (1) the group II introns found in both eubacterial and archaeal genomes, (2) retrons, which are also found in eubacteria and some archaean, and (3) the diversity generating retroelements that have thus far been found in the eubacteria.

2. GROUP II INTRONS

Among the best understood retroelements of prokaryotic organisms are the group II introns. The group II introns were initially discovered in the genomes of organelles like chloroplasts and the mitochondria of yeast (Michel and Land, 1985). Using degenerate primers and a PCR screening method, group II introns were subsequently discovered in a variety of different bacteria as well (Ferat and Michel, 1993). Like any intron DNA, group II introns interrupt a gene (although this occurs only occasionally in bacteria) and must be spliced out of the corresponding RNA transcript to yield a mature RNA encoding a functional gene. In the initial RNA transcript, independent of any host cell splicing factor, group II introns form an autocatalytic RNA that can self-splice from the RNA molecule. The excised intron RNA forms a characteristic lariat structure by a mechanism similar to the splicing of nuclear introns by the eukaryotic spliceosome (Saldanha et al., 1993).

In addition to self-splicing, most group II intron DNAs found in bacteria contain an open reading frame (ORF). The product of this ORF gene is a multifunctional protein with (1) maturase activity that aids in intron splicing (Matsurra et al., 2001), (2) RT activity that converts the spliced intron RNA into a cDNA copy and, in some cases (3) a DNA endonuclease that helps the intron insert into a new site in a DNA molecule (Martinez-Abarca and Toro, 2006; Zimmerly et al., 2001). This intron