Genomic libraries

Cloning DNA, by whatever method, gives rise to a population of recombinant DNA molecules, often in plasmid or phage vectors, maintained either in bacterial cells or as phage particles. A collection of independent clones is termed a **clone bank** or **library**. The term **genomic library** is often used to describe a set of clones representing the entire genome of an organism, and the production of such a library is usually the first step in isolating a DNA sequence from an organism's genome. A genomic library is a rich resource for the scientist, as it represents the entire genome of an organism and (at least in theory) should contain all the genes and their control sequences. The first consideration in constructing a genomic library is the number of clones required. This depends on a variety of factors, the most obvious one being the size of the genome. Thus, a small genome such as that of *E. coli* will require fewer clones than a more complex one such as the human genome. The type of vector to be used also has to be considered, which will determine size of fragments that can be cloned. In practice, library size can be calculated quite simply on the basis of the probability of a particular sequence being represented in the library. There is a formula that takes account of all the factors and produces a 'number of clones' value. The formula is:

$N = \ln(1 - P) / \ln(1 - a/b)$

where N is the number of clones required, P is the desired probability of a particular sequence being represented (typically set at 0.95 or 0.99), a is the average size of the DNA fragments to be cloned, and b is the size of the genome (expressed in the same units as a).

		No. clones N, $P = 0.95$	
Organism	Genome size (kb)	20 kb inserts	45 kb inserts
Escherichia coli (bacterium)	4.0×10^3	6.0×10^2	2.7×10^2
Saccharomyces cerevisiae (yeast)	1.4×10^4	2.1×10^3	9.3×10^2
Arabidopsis thaliana (simple higher plant)	7.0×10^4	1.1×10^4	4.7×10^3
Drosophila melanogaster (fruit fly)	1.7×10^5	2.5×10^4	1.1×10^4

Note: The number of clones (N) required for a probability (P) of 95% that a given sequence is represented in a genomic library is shown for a range of different organisms. Approximate genome sizes of the organisms are given (haploid genome size, if appropriate). Two values of N are shown, for 20 kb inserts (_ replacement vector size) and 45 kb inserts (cosmid vectors). The values should be considered as minimum estimates, as strictly speaking the calculation assumes (1) that the genome size is known accurately, (2) that the DNA is fragmented in a totally random

manner for cloning, (3) that each recombinant DNA molecule will give rise to a single clone, (4) that the efficiency of cloning is the same for all fragments, and (5) that diploid organisms are homozygous for all loci. These assumptions are usually not all valid for a given experiment

Gene expression

As shown in Fig. 2.2, **the flow of genetic information is from DNA to protein. Whilst a detailed knowledge of gene expression** .

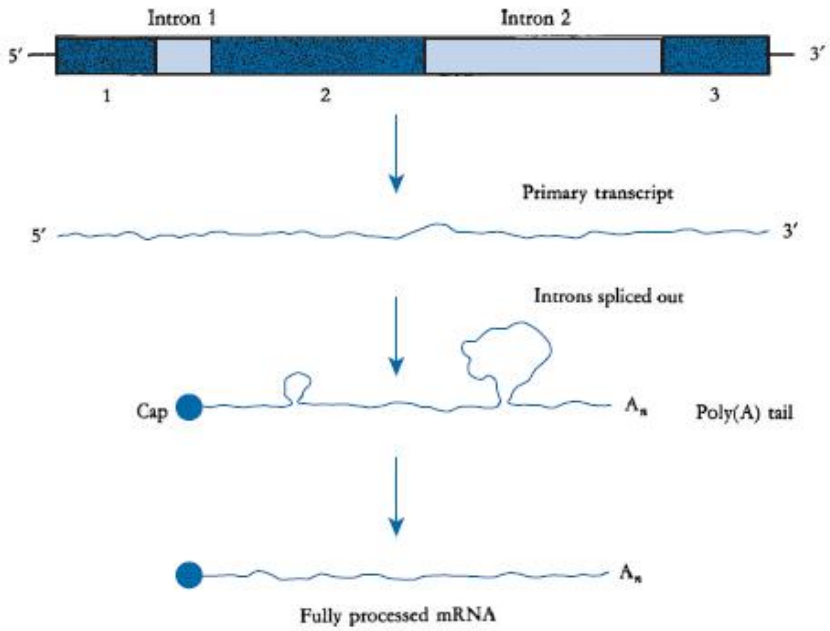

Fig. 2.8 Structure and expression of the mammalian ß-globin gene. The gene contains two intervening sequences or introns. The expressed sequences (exons) are shaded and numbered. The primary transcript is processed by capping, polyadenylation, and splicing to yield the fully functional mRNA.

Transcription and translation ,these two processes are the critical steps involved in producing functional proteins in the cell. **Transcription involves synthesis of an RNA from the DNA template provided by the non-coding strand of the transcriptional unit in question.** The enzyme responsible is **RNA** . In prokaryotes there is a single RNA polymerase enzyme, but in eukaryotes there are three types of RNA polymerase (I, II, and III). These synthesise ribosomal, messenger, and transfer/5 S ribosomal RNAs, respectively. **Transcription has several component stages: (1) DNA/RNA polymerase binding, (2) chain initiation, (3) chain elongation, and (4) chain termination and release of the RNA.** Promoter structure is important in determining the binding of RNA polymerase but will not be dealt with here. When the RNA molecule is released, it may be immediately available for translation (as in prokaryotes) or it may be processed and exported to the cytoplasm (as in eukaryotes) before translation occurs. Translation requires an mRNA molecule, a supply of charged tRNAs (tRNA molecules with their associated amino acid residues), and ribosomes (composed of rRNA and ribosomal

proteins). The ribosomes are the sites where protein synthesis occurs.The ribosome is a complex structure that essentially acts as a 'jig' that holds the mRNA in place so that the codons may be
matched up with the appropriate **anticodon** on the tRNA, thus ensuring that the correct amino acid is inserted into the growing polypeptide chain. The mRNA molecule is translated in a 5_→3_ direction,mcorresponding to polypeptide elongation from N terminus to C terminus. Although transcription and translation are complex processes, the may be summarised as shown in **Fig. 2.9.**

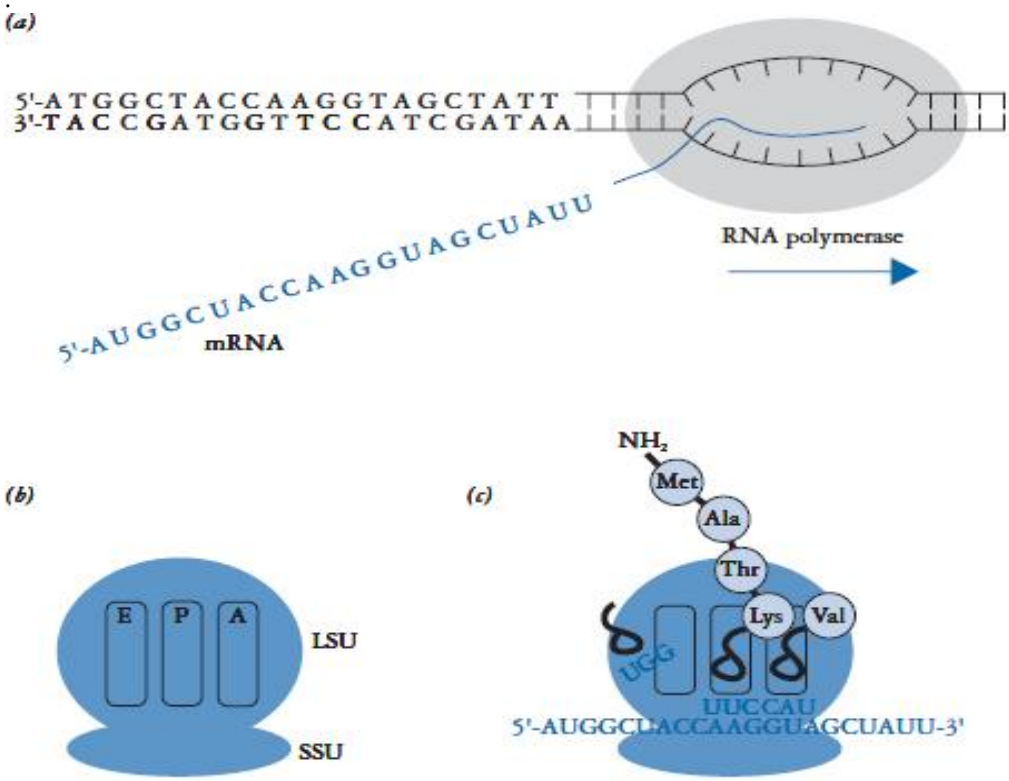

Fig. 2.9 Transcription and translation. (*a*) Transcription involves synthesis of mRNA by RNA polymerase. Part of the DNA/mRNA sequence is given. The mRNA has the same sequence as the coding strand in the DNA (the non-template strand), apart from U being substituted for T. (*b*) The ribosome is the site of translation and is made up of the large subunit (LSU) and the small subunit (SSU), each made up of ribosomal RNA molecules and many different proteins. There are three sites within the ribosome. The A (aminoacyl) and P (peptidyl) sites are involved in insertion of the correct tRNA–amino acid complex in the growing polypeptide chain. The E (exit) site facilitates the release of the tRNA after peptide bond formation has removed its amino acid. (*c*) The mRNA is being translated. The amino acid residue is inserted into the protein in response to the codon/anticodon recognition event in the ribosome. The first amino acid residue is encoded by AUG in the mRNA (tRNA anticodon TAC), which specifies methionine (see Table 2.1 for the genetic code). The remainder of the sequence is translated in a similar way. The ribosome translates the mRNA in a 5′→3′ direction, with the polypeptide growing from its N terminus. The residues in the polypeptide chain are joined together by peptide bonds.