# Chapter 7

## Web Text Box 1

### *In silico* cloning

As DNA sequencing became more efficient during the 1990's, databases were created to store and analyze this wealth of new genetic data. Today, the DNA sequences of many thousands of genes, and the protein sequences they encode, are readily available to researchers. For many organisms, entire genome sequences are accessible on line. In most cases, these mark the efforts of multinational scientific consortia, such as the Human Genome Project. The first genomes sequenced were of 'model' organisms such yeast and Drosophila. Later, these were joined by the genomes of many bacteria, domestic animals and crop plants, which were selected for detailed analysis because of their enormous economic potential. In addition to such genomic information there also exist massive databases of expressed sequence tags (EST). These are short sequences of complementary DNA (cDNA) that are generated from RNA by the enzyme reverse transcriptase. The RNA in turn is isolated from a particular tissue of a given organism or a particular developmental stage or from a particular type of tumor. Thus, these sequences represent only those genes that are expressed (translated into RNA) in that particular context e.g. in that type of tumor. Both whole genome and EST sequence resources are widely used in the identification of novel genes in a process referred to as *in silico* cloning. This can work in a number of ways. At its simplest, a scientist working in the laboratory produces a gene sequence, or even a fragment of sequence, from a particular organism but has no idea what the gene is or what it does. By pasting the sequence into a database our researcher can ask the computer whether similar sequences are present in the database. If the answer is 'no' then our researcher has to use his or her ingenuity to unravel the gene and its function. If, however, the answer is 'yes' and the database reports that, for example, the novel gene is >90% identical to the gene encoding glutamate dehydrogenase from yeast, *Drosophila* and *Homo sapiens* then a lot of tedious experimental work is saved.