# Chapter 14

# Gene Cloning and DNA Analysis in Medicine

*Chapter contents*

CHAPTER CONTENTS



Medicine has been and will continue to be a major beneficiary of the recombinant DNA revolution, and an entire book could be written on this topic. Later in this chapter we will see how recombinant DNA techniques are being used to identify genes responsible for inherited diseases and to devise new therapies for these disorders. First, we will continue the theme developed in the last chapter and examine the ways in which cloned genes are being used in the production of recombinant pharmaceuticals.

## 14.1 Production of recombinant pharmaceuticals

A number of human disorders can be traced to the absence or malfunction of a protein normally synthesized in the body. Most of these disorders can be treated by supplying the patient with the correct version of the protein, but for this to be possible the relevant protein must be available in relatively large amounts. If the defect can be corrected only by administering the human protein, then obtaining sufficient quantities will be a major problem unless donated blood can be used as the source. Animal proteins are therefore used whenever these are effective, but there are not many disorders that can be treated with animal proteins, and there is always the possibility of side effects such as an allergenic response.

We learned in Chapter 13 that gene cloning can be used to obtain large amounts of recombinant human proteins. How are these techniques being applied to the production of proteins for use as pharmaceuticals?

*Gene Cloning and DNA Analysis: An Introduction*. 6th edition. By T.A. Brown. Published 2010 by Blackwell Publishing.

### 14.1.1 Recombinant insulin

Insulin, synthesized by the β-cells of the islets of Langerhans in the pancreas, controls the level of glucose in the blood. An insulin deficiency manifests itself as diabetes mellitus, a complex of symptoms which may lead to death if untreated. Fortunately, many forms of diabetes can be alleviated by a continuing program of insulin injections, thereby supplementing the limited amount of hormone synthesized by the patient's pancreas. The insulin used in this treatment was originally obtained from the pancreas of pigs and cows slaughtered for meat production. Although animal insulin is generally satisfactory, problems may arise in its use to treat human diabetes. One problem is that the slight differences between the animal and the human proteins can lead to side effects in some patients. Another is that the purification procedures are difficult, and potentially dangerous contaminants cannot always be completely removed.

Insulin displays two features that facilitate its production by recombinant DNA techniques. The first is that the human protein is not modified after translation by the addition of sugar molecules (p. 236): recombinant insulin synthesized by a bacterium should therefore be active. The second advantage concerns the size of the molecule. Insulin is a relatively small protein, comprising two polypeptides, one of 21 amino acids (the A chain) and one of 30 amino acids (the B chain; Figure 14.1). In humans these chains are

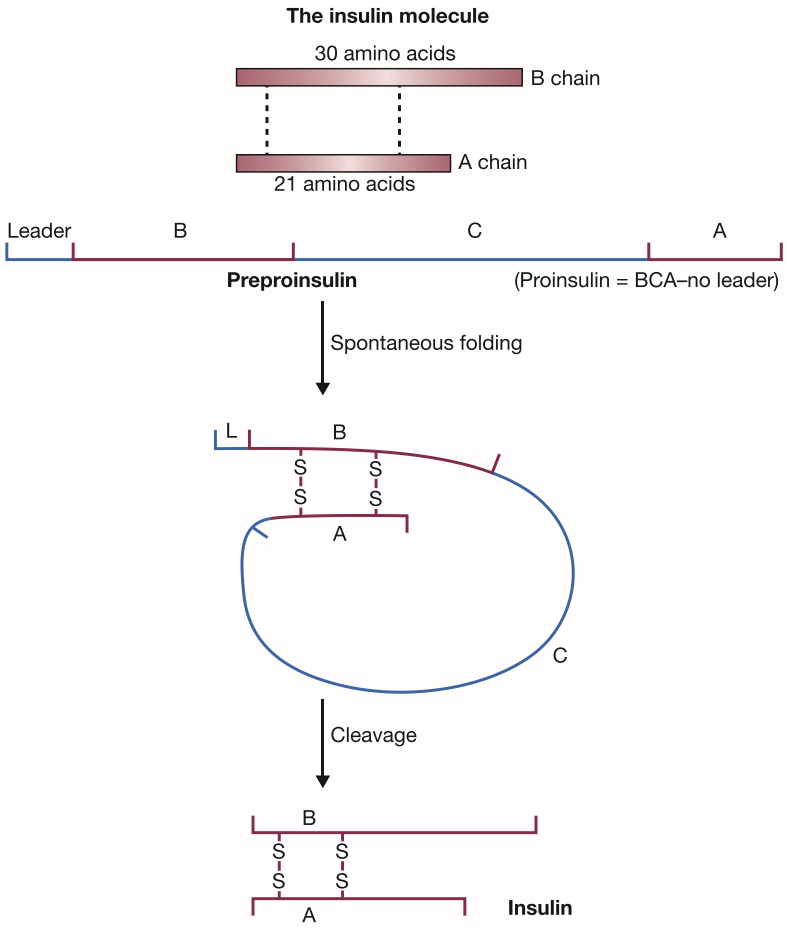



**Figure 14.1**

The structure of the insulin molecule and a summary of its synthesis by processing from preproinsulin.

synthesized as a precursor called preproinsulin, which contains the A and B segments linked by a third chain (C) and preceded by a leader sequence. The leader sequence is removed after translation and the C chain excised, leaving the A and B polypeptides linked to each other by two disulphide bonds.

Several strategies have been used to obtain recombinant insulin. One of the first projects, involving synthesis of artificial genes for the A and B chains followed by production of fusion proteins in *E. coli*, illustrates a number of the general techniques used in recombinant protein production.

#### *Synthesis and expression of artificial insulin genes*

In the late 1970s, the idea of making an artificial gene was extremely innovative. Oligonucleotide synthesis was in its infancy at that time, and the available methods for making artificial DNA molecules were much more cumbersome than the present-day automated techniques. Nevertheless, genes coding for the A and B chains of insulin were synthesized as early as 1978.

The procedure used was to synthesize trinucleotides representing all the possible codons and then join these together in the order dictated by the amino acid sequences of the A and B chains. The artificial genes would not necessarily have the same nucleotide sequences as the real gene segments coding for the A and B chains, but they would still specify the correct polypeptides. Two recombinant plasmids were constructed, one carrying the artificial gene for the A chain, and one the gene for the B chain.

In each case the artificial gene was ligated to a *lacZ′* reading frame present in a pBR322-type vector (Figure 14.2a). The insulin genes were therefore under the control of the strong *lac* promoter (p. 231), and were expressed as fusion proteins, consisting of the first few amino acids of β-galactosidase followed by the A or B polypeptides (Figure 14.2b). Each gene was designed so that its β-galactosidase and insulin segments were separated by a methionine residue, so that the insulin polypeptides could be cleaved from the β-galactosidase segments by treatment with cyanogen bromide (p. 233). The purified A and B chains were then attached to each other by disulphide bond formation in the test tube.

The final step, involving disulphide bond formation, is actually rather inefficient. A subsequent improvement was to synthesize not the individual A and B genes, but the entire proinsulin reading frame, specifying B chain–C chain–A chain (see Figure 14.1). Although this is a more daunting proposition in terms of DNA synthesis, the prohormone has the big advantage of folding spontaneously into the correct disulphide-bonded structure. The C chain segment can then be excised relatively easily by proteolytic cleavage.

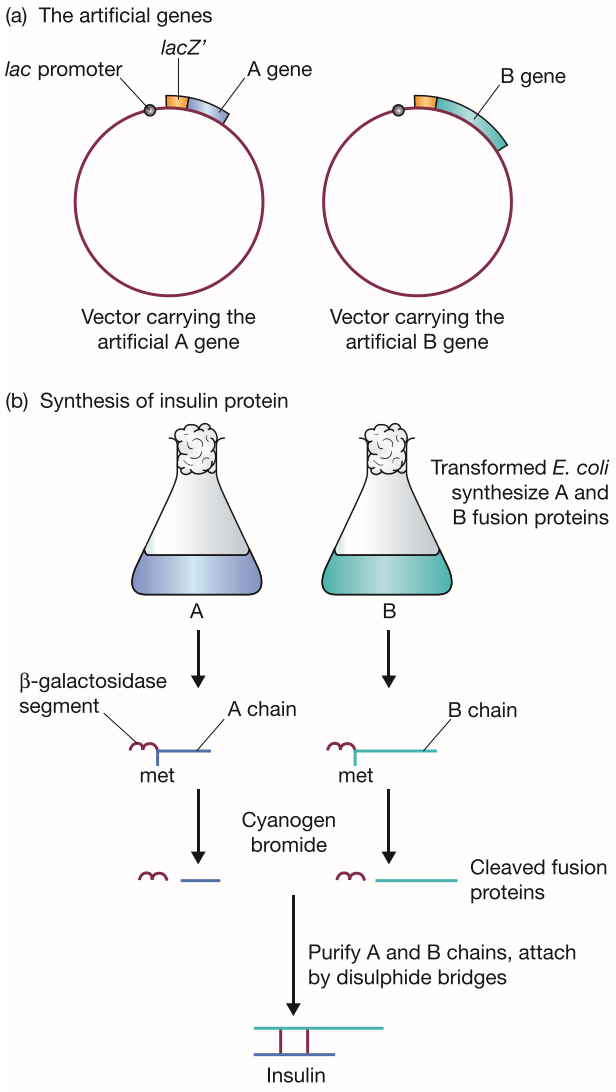
## 14.1.2 Synthesis of human growth hormones in E. coli

At about the same time that recombinant insulin was first being made in *E. coli*, other researchers were working on similar projects with the human growth hormones somatostatin and somatotrophin. These two proteins act in conjunction to control growth processes in the human body, their malfunction leading to painful and disabling disorders such as acromegaly (uncontrolled bone growth) and dwarfism.

Somatostatin was the first human protein to be synthesized in *E. coli*. Being a very short protein, only 14 amino acids in length, it was ideally suited for artificial gene synthesis. The strategy used was the same as described for recombinant insulin, involving insertion of the artificial gene into a *lacZ′* vector (Figure 14.3), synthesis of a fusion protein, and cleavage with cyanogen bromide.

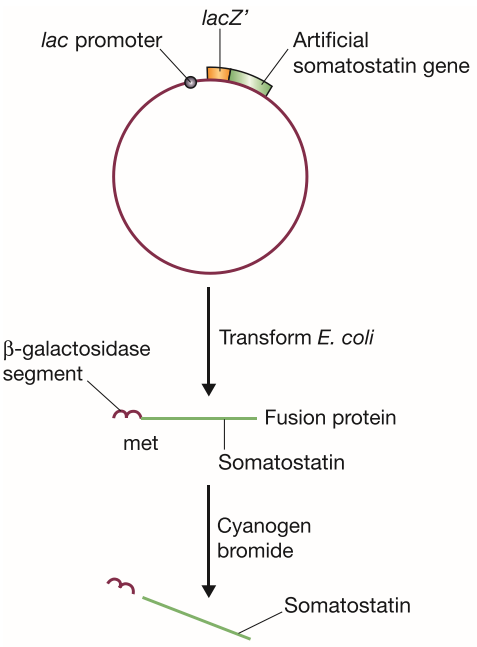
## Figure 14.2

The synthesis of recombinant insulin from artificial A and B chain genes.




## Figure 14.3

Production of recombinant somatostatin.



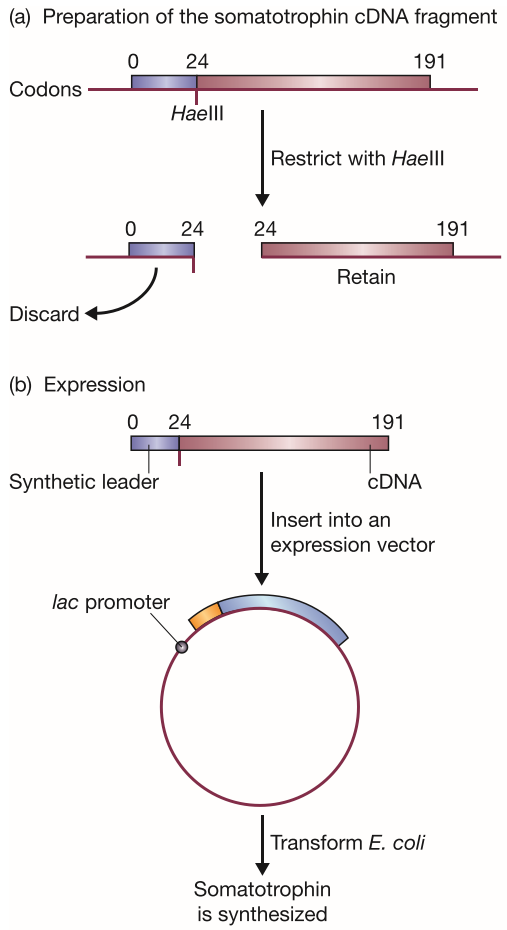



**Figure 14.4**
Production of recombinant somatotrophin.

Somatotrophin presented a more difficult problem. This protein is 191 amino acids in length, equivalent to almost 600 bp, an impossible prospect for the DNA synthesis capabilities of the late 1970s. In fact, a combination of artificial gene synthesis and complementary DNA (cDNA) cloning was used to obtain a somatotrophin-producing *E. coli* strain. Messenger RNA was obtained from the pituitary, the gland that produces somatotrophin in the human body, and a cDNA library prepared. The somatotrophin cDNA contained a single site for the restriction endonuclease *Hae*III, which therefore cuts the cDNA into two segments (Figure 14.4a). The longer segment, consisting of codons 24–191, was retained for use in construction of the recombinant plasmid. The smaller segment was replaced by an artificial DNA molecule that reproduced the start of the somatotrophin gene and provided the correct signals for translation in *E. coli* (Figure 14.4b). The modified gene was then ligated into an expression vector carrying the *lac* promoter.

## 14.1.3 Recombinant factor VIII

Although a number of important pharmaceutical compounds have been obtained from genes cloned in *E. coli*, the general problems associated with using bacteria to synthesize foreign proteins (p. 234) have led in many cases to these organisms being replaced by eukaryotes. An example of a recombinant pharmaceutical produced in eukaryotic cells is human factor VIII, a protein that plays a central role in blood clotting. The commonest form of hemophilia in humans results from an inability to synthesize factor VIII,