



VCE BIOLOGY 2006–2011

Introduction

There has been an ‘explosion’ of knowledge and understanding, and research in biology over the last 10 years. The successful completion of the Human Genome Project is one example of this rapid expansion where laboratory work and computer analysis operated in tandem to produce profound outcomes in a relatively short time span.

Contemporary studies in biology require students to develop an understanding of molecular biology including terms such as genome, proteome, bioinformatics and phenomics. Students should also be encouraged to consider the future possibilities of research and breakthroughs and any associated community, social or ethical issues.

To assist teachers to implement the VCE Biology Study Design 2006–2009 the following expert paper has been prepared to provide up-to-date information and explanation of important terms and concepts, and is of relevance to Units 3 and 4.

PROTEOMES, GENES AND JUNK DNA

By Stewart Jackel

DNA

In 1953 Edmund Hillary climbed Mt Everest and went home again and Elizabeth II was crowned Queen of Great Britain and Australia. At 24, tucked away in Cambridge University, James Watson was busy improving his tennis much to the annoyance of Francis Crick who wanted to get on with working out the structure of DNA. They succeeded and made the most significant scientific discovery since Mendel, or perhaps Darwin, beating the rival team of Maurice Wilkins and Rosalind Franklin in Oxford and Linus Pauling as the third race participant. It won them the 1962 Nobel Prize in medicine and physiology, which they shared with Maurice Wilkins. Rosalind Franklin, who had died from cancer in 1958, was largely overlooked in the credits.

The stuff of Nobel prizes is now the stuff of school science courses. And DNA, or more correctly, the ability of scientists to manipulate it, has become the stuff of science fiction.

DNA can be regarded as a polymer made up of a chain of 4 possible monomers joined to form a very long chain. Each DNA monomer has a backbone section that consists of a phosphate and a 5-carbon sugar-derived unit (deoxyribose). One of four possible bases containing nitrogen is attached to the deoxyribose sugar molecule. Each monomer, therefore, is a nucleotide comprising a phosphate, a pentose sugar and a base. The bases of DNA are thymine and cytosine (the pyrimidines) and adenine and guanine (the purines). The phosphate of one monomer can be joined to the sugar of the next to form a very long chain, the DNA molecule. DNA is always double stranded. The monomers in DNA bond across the molecule so that a thymine base always bonds with an adenine base while a cytosine base always bonds with a guanine base. The overall structure of the molecule is like a ladder: the alternating sugar and phosphate form the uprights and the bonded bases form the rungs. The ladder is twisted into a spiral to form a double helix.

The chemical bonding rules of the bases (C & G, A & T) ensures that, if the molecule is split along the middle, the rungs, a new half can be built by addition of the appropriate nucleotides. One side of a DNA molecule acts as a template for the other.

The order of the bases is an information code for the production of proteins. A sequence, or length of bases, specifies a sequence of amino acids – a polypeptide – the precursor of a protein.

Genes

A gene is a piece of information that specifies a protein. Structurally, a gene is a sequence of nucleotides – the 'active' constituents of DNA. Most cellular organisms have two chromosomes of each type and each type of chromosome has information at the same given point – the locus – for a gene. A chromosome pair can have a maximum of two alternatives – alleles – of a gene at each locus; the two alleles may be the same or different. This means that each locus may have information for one or two proteins.

Genome

An individual organism has a selection, at a given locus, of the possible alleles available from those present in the population – the gene pool. The selection possessed by one individual makes up its genetic complement. The term 'genome' can refer to the genetic complement of an organism or the total alleles shown by a species. The Human Genome Project or the Tamar Wallaby (*Macropus eugenii*) Genome Project are two examples. In both cases the entire genome – the total gene sequence – for the species has been identified so that all possible alleles have been mapped on their respective chromosomes. These are both huge enterprises: the Tamar wallaby has 3 billion base pairs in its genome (ARC Centre for Kangaroo Genomics, <http://kangaroo.genomics.org.au>).

There is no relationship, though, between the complexity of an organism and the size of its genome. Humans have a genome that is smaller than some single-celled organisms. This is the C-value enigma.

Coding and Non-coding DNA

Not all the DNA codes for proteins. Some sequences appear to code for control mechanisms of the coding sequences, others appear to code for nothing. The total nucleotide sequence of a chromosome comprises coding and non-coding sequences.

Transcription

Nuclear DNA does not leave the nucleus. The code is copied onto RNA and it is the RNA copy that takes the code from the nuclear store to the site in the cytoplasm – the ribosomes – where it is expressed.

RNA

In order for the DNA code to arrive at the organelles in the cytoplasm where polypeptides are assembled, the information must be transcribed – copied – (from the Latin *transcibere*, to copy off) and then translated – relocated – (from the Latin *translatius*, carried over). In short, the information is copied then the copy is moved.

Many types of RNA exist, but two types of RNA are involved in the sequential processes of transcription and translation: mRNA and tRNA.

RNA is a polymer with a similar skeleton to that of DNA; a backbone section that consists of a phosphate and a 5-carbon sugar (ribose) unit. One of four possible bases containing nitrogen is attached to the ribose molecule. Each monomer, therefore, is a nucleotide comprising a phosphate, a pentose sugar and a base. The bases of RNA are uracil and cytosine (the pyrimidines) and adenine and guanine (the purines). Most RNA is single-stranded.

The enzyme RNA polymerase promotes the production of a molecule of mRNA which is complementary to the section of DNA involved. The process is partially proof read but the control system is less effective than it is for the process of DNA replication.

mRNA

Each messenger RNA, mRNA, is a complementary copy of a gene – a specific section of DNA. The initial stages of transcription produce mRNA molecules copied from most of the length of the gene concerned. The non-coding sequences – the introns – are spliced out and the remaining coding sequences – the exons – are ligated together to form the complete molecule. Some primary transcripts contain up to several million bases that are reduced to several tens of thousand bases once the introns are removed.

The mRNAs are translocated to ribosomes in the cytoplasm where they code for the synthesis of the specific amino acid sequence that will form the polypeptide and hence the functioning protein.

Some gene mutations cause errors in the mRNA splicing process by causing the omission of some introns. The errors can produce dysfunctional proteins that do not function normally so that some cause disease. It is now theoretically possible to correct the splicing errors and so remove the cause of some diseases such as some cancers.

Translation

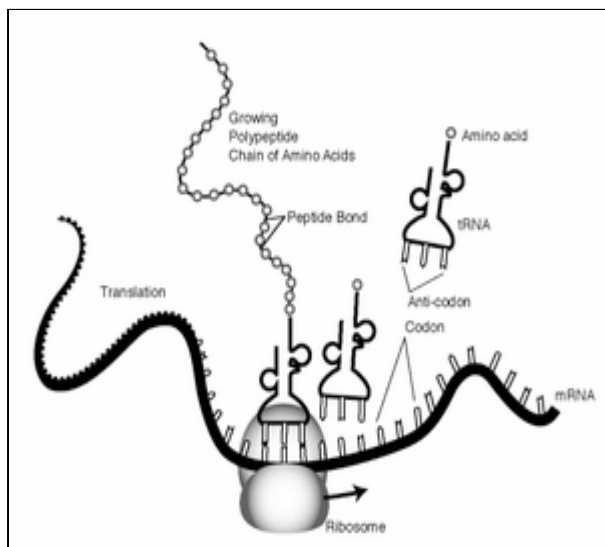
Transfer RNA, tRNA, has similar chemistry to that of mRNA: it is a single-stranded polymer of the same units as mRNA but it is much smaller and has three bases (an anticodon).

The process – the production of polypeptides – takes place on ribosomes, or molecular work-benches. The mRNA is 'read' in blocks of 3 bases – the codons – where each codon specifies a particular amino acid. mRNA is a complementary copy of the DNA, that is, each mRNA codon has information complementary to that of the gene from which it is formed.

Translation begins when a sub-unit of the ribosome bonds with the 'start' codon – usually AUG – on the mRNA.

Types of tRNA molecules, each with one of the 20 specific amino acids attached, are available in the cytoplasm. Different types of tRNA are characterised by a specific anticodon which is complementary to the mRNA codons.

As each mRNA codon along the length of the transcribed and relocated mRNA is exposed on the ribosome a complementary tRNA with its attached amino acid temporarily bonds to the mRNA. The adjacent amino acids bond form peptide bonds to form a growing chain – a polypeptide – the primary structure of a protein.



'Translation (genetics)', Wikipedia

http://en.wikipedia.org/wiki/Translation_%28genetics%29

Polypeptides

The synthesised polypeptide is released from the ribosome and the used tRNA molecules move back away from the ribosome to bond to another amino acid of the specific type. Most polypeptides must be re-structured in order to perform their designated function – as enzymes, hormones or cell structural components.

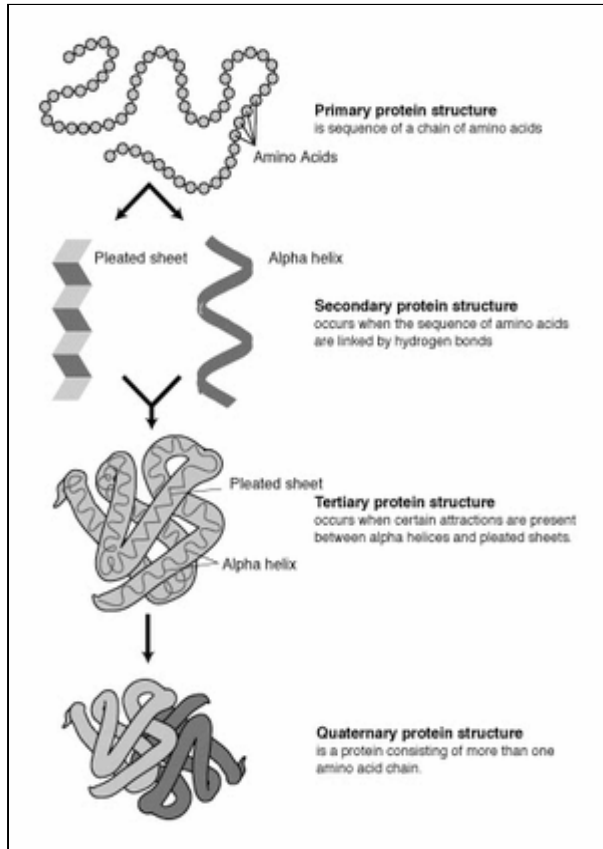
Proteins

Each step in each biochemical reaction in living things is controlled by a specific enzyme. Each enzyme is a protein and each protein is constructed from one or more polypeptides. The amino acid sequence of a polypeptide is determined by the mRNA codon sequence which is determined by the DNA base sequence of which it is a complementary copy.

Polypeptides fold and are held in place by hydrogen bonds to form the secondary structure of the protein. The tertiary structure is held in place by hydrogen, ionic and disulphide bonds. A quaternary structure is formed by the bonding of two proteins. Generally the shape of the protein determines its enzymatic or structural function.

Conventional wisdom is that one gene specifies one protein. However the 'truism' is clouded by the observation that, while a human has 3.5 billion base pairs in its genome, only about 25000 genes have been identified to code for the 400000 proteins in the human proteome.

It seems that the cellular environment of the cell polypeptide production has an influence on the structure of the protein formed, so that one gene can code for more than one protein, depending on its environment, that is, its location in the organism.



'Protein Structure', Wikipedia
http://en.wikipedia.org/wiki/Protein_structure

Proteomes

The entire range of genes of an organism (or a species) comprises its genome. Since the genes specify the organism's proteins, the genome specifies the proteome – the entire range of proteins of an organism (or a species).

Other RNAs

It seems that many types of RNA other than mRNA and tRNA are important in living things. Molecules termed small RNAs are not involved in the transcription–translation process but regulate mRNA molecules that are. They bind to specific targets, which are then degraded rather than translated into protein. Small RNAs regulate genes. They may be associated with biological processes, such as responses to stress.

It has recently been discovered that some RNA is double stranded and may be up to 200 nucleotides in length. These dsRNAs have been used to silence targeted genes in a range of cell types in plants and animals. The dsRNAs begin the RNA interference, RNAi, pathway. The molecules are cut into smaller dsRNAs about 20–25 nucleotides long to produce small interfering RNAs, siRNAs. The siRNAs assemble into complexes that contain endoribonuclease which are RNA-induced silencing complexes, RISCs, and unwind in the process. The siRNA strands then guide the RISCs to complementary RNA molecules, where they cleave and destroy RNA.

Junk DNA

About 3% of human DNA codes for proteins. The remaining 97% is said to be junk, perhaps a left-over from evolutionary history. Much of the junk DNA consists of multiple repeats of lengths of DNA that appear to serve no useful function other than that they provide indicators of human phylogeny.

It has recently been suggested that there may be a relationship between the introns and proteins and that the introns influence the expression of an exon. One investigation involved modification of a non-coding sequence DNA of a plant. The change caused a change in the leaf structure, suggesting a causal relationship. However the removal of 1% of a mouse genome appeared to have no detectable effect.

It has been suggested that:

1. Much of the DNA may, in fact have no function.
2. Non-coding regions may be remnants from the organism's phylogeny – genes that have been fractured over time.
3. Some non-coding regions may act as spacers to allow more room for enzymes to interact with the coding DNA.
4. The non-coding DNA may be evidence of past retroviral infections.
5. We may simply have not found the use yet, such as regulating gene expression.
6. Like flocking behaviour in birds, an individual coding sequence gene is statistically less likely to be damaged by chemicals, radiation or crossing over so non-coding DNA protects the coding regions.
7. Non-coding regions may be a reservoir of raw material for evolution to act on.

References

Cold Spring Harbor Laboratory 2003, 'Designer molecules correct RNA splicing defects', BrightSurf.com The resource for science information
http://www.brightsurf.com/news/jan_03/CSHL_news_011503.html

Public Library of Science 2003, 'MicroRNA targets: How big is the iceberg?', BrightSurf.com The resource for science information
http://www.brightsurf.com/news/oct_03/PLS_news_101403.php