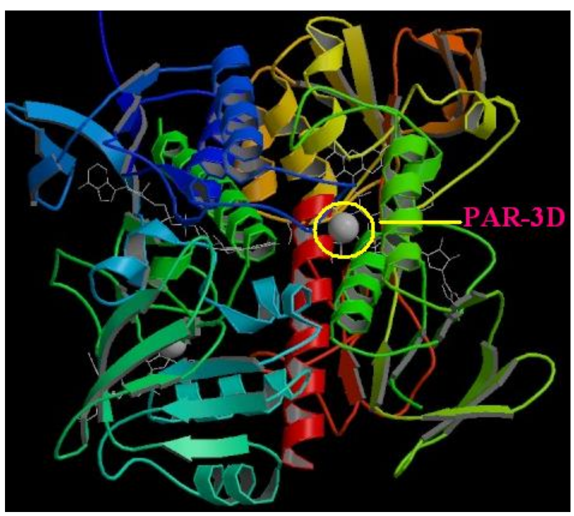
Libby MacKinnon
CSE 527 notes
October 1, 2007
Intro to Molecular Biology Continued

Proteins

Proteins are chains of amino acids. They range in length from hundreds to tens of thousands of amino acids. There are 20 different kinds of standard amino acids made and used in cells. Proteins are the major functional elements in cells. They make up the major structural and mechanical elements, enzymes which catalyze reactions, and receptors for hormones and other signaling molecules. They are not only the product of the transcription and translation process, but also can be transcription factors, controlling the rate of gene expression.

Proteins do not remain in their long chain form. The complex folding process they undergo to reach their 3D structure is crucial to their function, and a lot of research is done to understand protein folding. The actual active site of the protein makes up only a fraction of the overall size, as seen in the example of 3D structure below.

Figure 1: The active site of the protein is shown along with its 3D structure. http://sunserver.cdfd.org



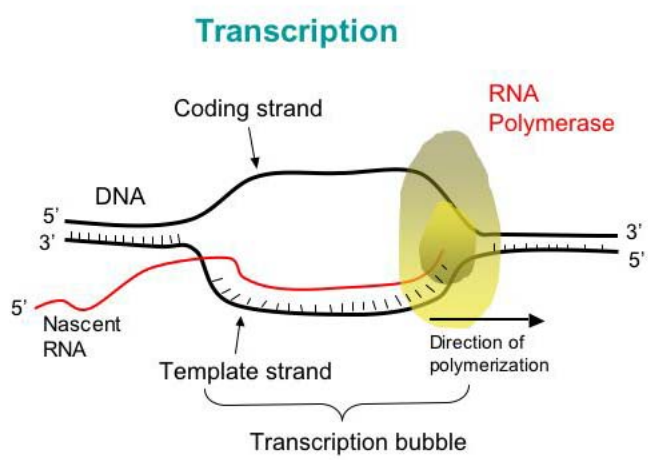
Central Dogma

The "central dogma" of molecular biology (a term coined by Francis Crick 50 years ago) describes the path from gene to protein. First the DNA of the gene in question is transcribed into a segment of messenger RNA (mRNA). This mRNA is translated separately into a protein. Each step is complicated. Since the notion of the central dogma came about, there have been many instances of deviation from this natural flow of information, but the idea is still critical to the study of genetics.

Transcription

Transcription is the copying of a piece of DNA by RNA polymerase, resulting in a strand of mRNA that is a duplicate of the ”sense” (or “coding”) strand of the original DNA. DNA is inherently easy to copy. The antisense strand serves as a natural template for duplication and a copy of the sense strand is achieved by simply matching up the correct nucleic acids to this template.

Figure 2: An mRNA strand is made from a DNA template by RNA polymerase. http://www.langara.bc.ca/biology/mario/Assets/transcription.jpg



Translation

Translation is the process of making a protein from the mRNA strand that was transcribed. Because the mRNA is made up of 4 nucleic acids (A,C,G,U) and proteins are made up of 20 amino acids, there has to be some way to determine how the protein chain is made up. This is called the genetic code. Any 3-nucleic acid sequence (codon) codes for a certain amino acid (or for the synthesis to stop).

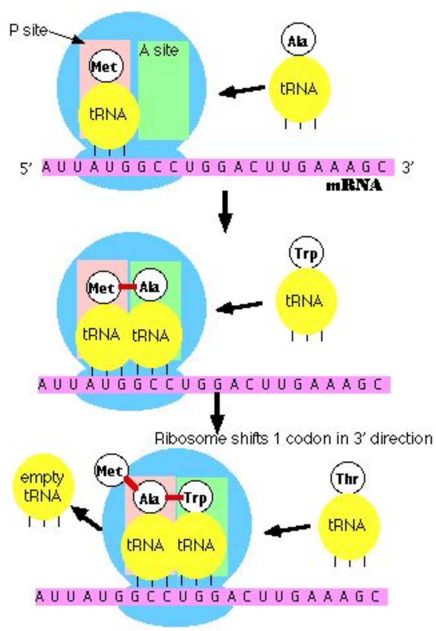
Figure 3: Each three letter sequence of nucleic acids in RNA codes for a certain amino acid or action in translation. http://www.clcbio.com/scienceimages/

Second base in codon

First base in codon	U	C	A	G	Third base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

A ribosome reads from the start codon (usually AUG), and keeps adding the amino acid that is called for at each codon to the growing chain. Then the ribosome moves forward 3 positions on the mRNA. In some cases, especially in bacteria, several different proteins can be encoded one after the other on a single mRNA, and many copies of each can be synthesized at once from the same strand of mRNA, as multiple ribosomes traverse it.

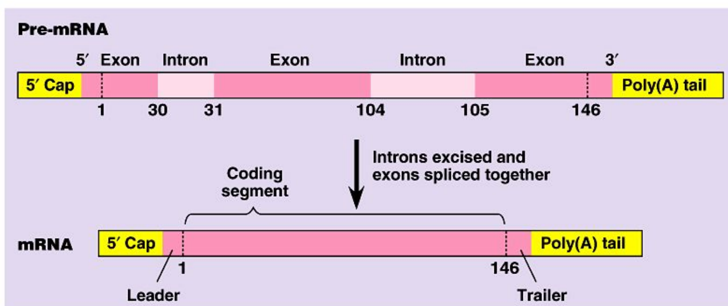
Figure 4: This is the basic process of the ribosome "reading" the codon and attaching the amino acid chain to the next amino acid, with the help of tRNA. http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/T/Translation.gif



Gene Structure

DNA and RNA are always synthesized from the 5' to the 3' end. Genes are described as being transcribed from 5' to 3' also, i.e., 5' to 3' with respect to the sense strand. There is usually a promoter region before the 5' end that signals where to begin copying. At each end of the transcribed strand there are segments that are untranslated, and in Eukaryotes a long string of A's is attached to the 3' end (poly(A) tail; this is not directly coded in the DNA). Also, in most eukaryotes there are many regions called introns, that are spliced out between transcription and translation. In fact, only a small percentage (less than 2%) of the original gene is actually expressed.

Figure 5: A simplified view of gene structure, showing the expressed and non-expressed regions. http://kentsimmons.uwinnipeg.ca/cm1504/



Genome Sizes

The genome is all the hereditary information of every cell of an organism. It is the entirety of the DNA. It is important to note that the size of the genome does not necessarily correlate with the complexity of the organism it belongs to. For example, the lungfish has a larger genome than humans, by two orders of magnitude. Several examples of known genome sizes are shown below.

Genome Surprises

With the exploration of the human (and other) genome, came many surprises. Humans have less than a third as many genes as were expected. However, from those genes came more proteins than expected due to alternative splicing, alternative starts, and alternative tails of the genes.

Protein-wise, all mammals are nearly the same, but more variation that expected among individuals was discovered.

Figure 6: The genome sizes of several organisms are given, from class notes.

	Base Pairs	Genes
Mycoplasma genitalium	580,073	483
MimiVirus	1,200,000	1,260
E. coli	4,639,221	4,290
Saccharomyces cerevisiae	12,495,682	5,726
Caenorhabditis elegans	95,500,000	19,820
Arabidopsis thaliana	115,409,949	25,498
Drosophila melanogaster	122,653,977	13,472
Humans	3.3×10^9	~25,000

It was also determined that there is probably more non-coding RNAs than coding, and many of the non-coding regions are common across all vertebrates. While 90% of DNA is transcribed, less than 2% is coding.

Also, there is "epigenetic" information that is inherited but does not come from DNA. Examples of this are cell differentiation and histones, proteins that package DNA and can promote or repress transcription.