DNA to Protein Overview
Primary Knowledge (PK)
Participant Guide

Description and Estimated Time to Complete

This learning module provides information needed to understand how the digitally encoded information in DNA is translated into a functional protein that can be used for biomedical applications such as diagnostics, analysis and measurements. Activities delve deeper into protein structure and function as well as gene transcription.

The PK unit is your reading material and an overview of the DNA to Protein concepts. This information will help you to better understand the applications of MEMS (MicroElectroMechanical Systems) within the medical field. This is the basic knowledge needed to research and understand protein structure and function. This information will also be used to transcribe and translate a gene in the related activity.

Estimated Time to Complete
Allow approximately 15 minutes.

Introduction

Most of the properties of living organisms ultimately arise from a class of molecules known as proteins. Proteins are polymers composed of subunits known as amino acids. These linear polymers fold into specific three-dimensional structures with specific, unique functions. Amino acids dictate the structure of the protein. The linear sequence of information found within a gene in an organism's DNA dictates the order of amino acids.
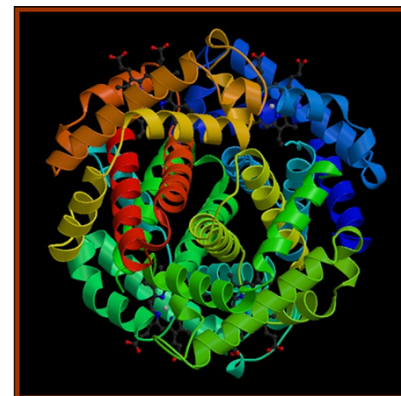
Protein Tertiary Structure
[Image courtesy of RCSB
Protein Data Bank]

Many applications take advantage of the functions of proteins. Here are a couple of applications that are used today:
- Diabetics rely on glucometer strips that contain specific protein enzymes capable of catalyzing reactions that aid in determining blood glucose levels.
- Home pregnancy tests rely on proteins called antibodies found in the urine of pregnant women. These proteins react with the coating on the pregnancy test strip enabling a positive or negative result. *(See graphic below)*
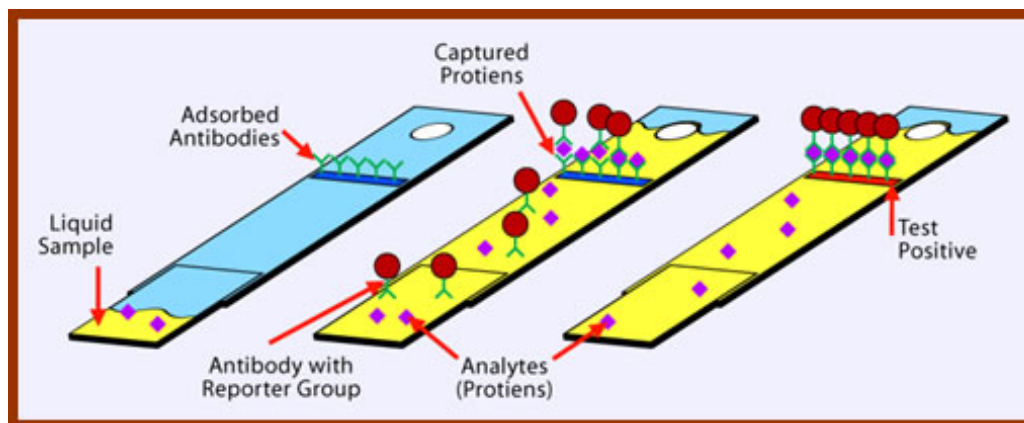
Home Pregnancy Test

MEMS Applications within Proteomics

The emerging field of proteomics is analyzing the complex range of protein expression within cells. Biologists use this information to create protein microarrays on chips. These arrays investigate both the interactions of the arrayed proteins with other proteins as well as their potential for chemical modification. These protein microarrays and their electronic interface are microelectromechanical systems or MEMS, or more specifically, bioMEMS.

Another bioMEMS application in proteomics is ChIP/chip analysis. This bioMEMS provides researchers the ability to make global measurements of protein/DNA interaction.

These methodologies rely on an understanding of the way in which information encoded digitally in DNA translates into a functional protein. This unit helps to answer these questions:

- How did researchers crack this genetic code?
- Is the genetic code universal?
- What do we know about information flow within a cell?

Objectives

- Describe the genetic code.
- Explain the "central dogma" of biology: DNA → RNA → Protein (polypeptide), and differentiate the separate parts that comprise the dogma.
- Describe the function of the components of Protein Translation.

Key Terms (Definitions at end of guide)

Alkaptonuria
Amino acid
Antibody
Anticodon
ChIP/chip analysis
Codon
Enzyme
Ferment
Genetic code
Promoter
Protein
Proteome
Proteomics
Ribonucleotide
Ribosome
Template
Translation
Transcription
Transcription factor
Wobble

The Central Dogma of Biology

Studies that provided clues to the idea for "The Central Dogma of Biology":

In 1902, Archibald Garrod, an English physician, treated an infant whose diapers had a dark reddish black stain. He recognized this stain as a rare condition named alkaptonuria. The urine of patients with this condition turns black upon exposure to air.

Garrod and his colleagues did not stop at the diagnosis. They proposed that the infant was missing a "certain ferment" (enzyme) that the body would normally use to break down the alkapton prior to excretion. They also knew that alkaptonuria was a recessive genetic disorder. Substituting enzyme for "ferment", Garrod and his colleagues described a connection between a gene and a protein.

In the 1940's, George Beadle and Edward Tatum exposed the spores of the fungi Neurospora to x-rays, and isolated the mutant progeny. They concluded that one gene specifies one enzyme.

Based on these and other previous findings, in 1957 Francis Crick proposed that genetic information flows in one-direction, from DNA to RNA to protein. Crick named this proposal the "Central Dogma" of biology.

Today, the Central Dogma is often revised to *DNA is transcribed to RNA*, and *RNA is translated to a polypeptide (see graphic)*. This revision takes into account functional proteins that are composed of more than a single polypeptide. Exceptions do occur, and one major exception to this flow of information is in the family of Retroviruses. These viruses encode a polymerase, reverse transcriptase, which uses a RNA template to make a DNA copy.
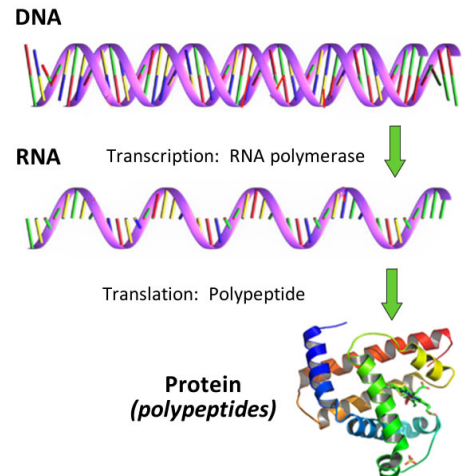
The RNA Molecule

RNA is a similar molecule to DNA with some very specific differences. Both RNA and DNA are long polymeric molecules with a ladder backbone and rungs of nitrogenous bases. In RNA, the sugar in the ladder is ribose rather than deoxyribose (as in DNA), and the nitrogenous base uracil replaces the thymine found in DNA. RNA is single-stranded and DNA is double-stranded as shown in the graphic.
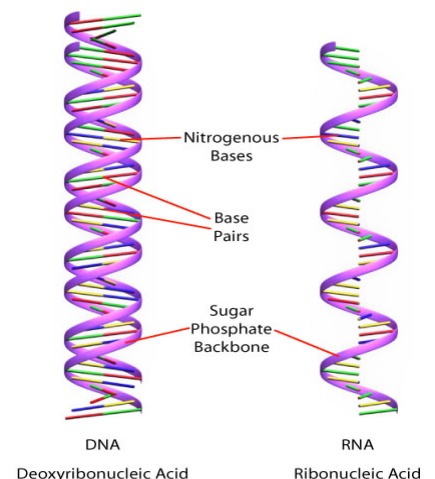
RNA's Role in "DNA to Protein"

At least three types of RNA are known to be involved in the information flow from DNA to protein: mRNA, tRNA and ribosomal RNA. We will identify these RNA as we move through the Central Dogma (DNA to RNA to Protein).

The first step of DNA to RNA is transcription. Transcription is defined as DNA-directed RNA synthesis. It requires a DNA template, RNA polymerase (*an enzyme that produces RNA*), and ribonucleotide subunits. Transcription occurs in three phases: initiation, elongation and termination.

Transcription: RNA polymerase

DNA

Initiation
Elongation
Termination

mRNA
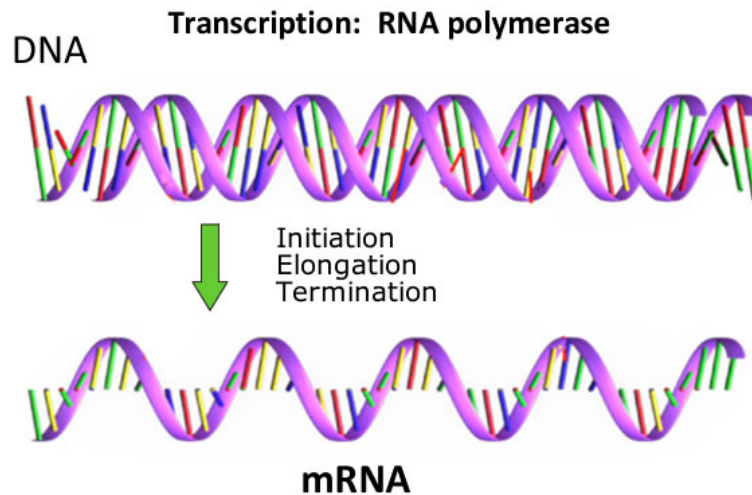
Initiation requires a promoter and RNA polymerase. The promoter contains the initiation site and tells RNA polymerase where to start, which strand to read, and which direction to take from the start. RNA polymerase then elongates the transcript until termination occurs at specific base sequences.

Transcription produces a RNA complementary to the DNA. This RNA is called the messenger RNA (**mRNA**). In eukaryotic cells, the mRNA leaves the nucleus and enters the cytoplasm *(see graphic)*. In the cytoplasm, the mRNA serves as the template for the creation of a polypeptide (protein).
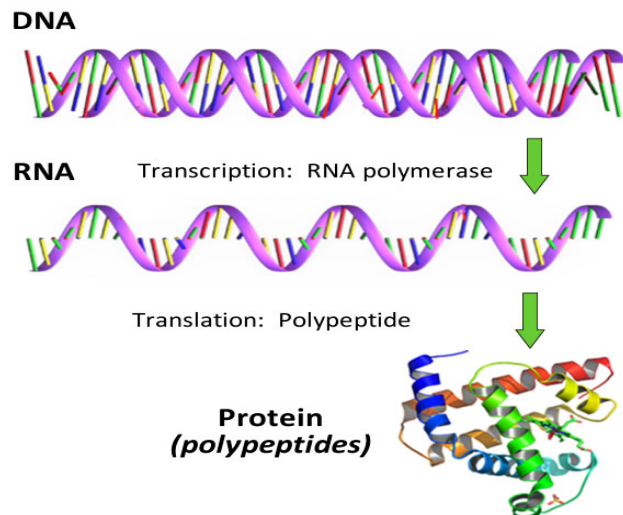
Eukaryotic Cell

Cytoplasm

Nucleus
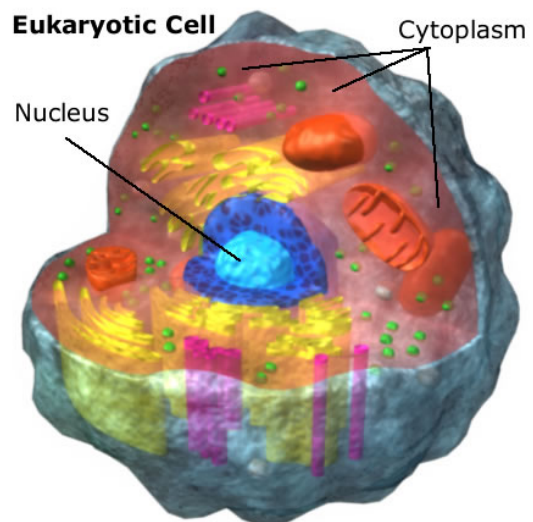
This second step of the Central Dogma, RNA to protein, is known as translation where the mRNA is translated to protein or polypeptide.

Translation: Polypeptide

mRNA

Protein
(polypeptides)
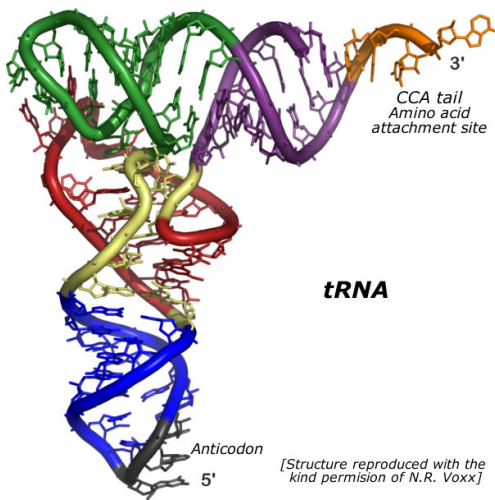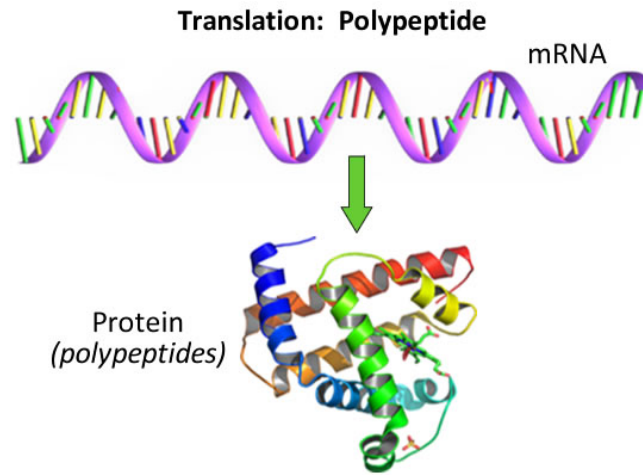
During translation another type of RNA, transfer RNA (tRNA), acts as an adapter molecule that recognizes the message of mRNA. The tRNA carries specific amino acids for incorporation into a growing polypeptide chain.

A third RNA, ribosomal RNA (rRNA) is part of the ribosome that is also essential for translation of the mRNA. (There are other RNAs of increasing interest to molecular biologists, and these small RNAs function in the control of plant and animal gene expression. These small RNAs are not part of this discussion.)

3'
CCA tail
Amino acid
attachment site

tRNA

Anticodon

5'

[Structure reproduced with the kind permision of N.R. Voxx]

In summary…
The Central Dogma of Biology

DNA

RNA   Transcription: RNA polymerase

Translation: Polypeptide
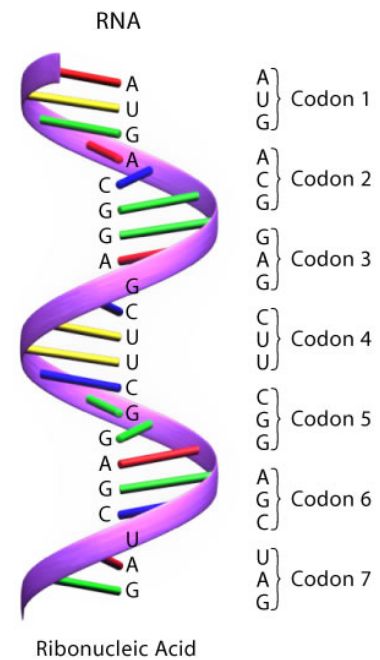
Protein
(polypeptides)

The Genetic Code

Although the flow of information was postulated in 1957 via the Central Dogma of Biology, the genetic code of DNA was not deciphered until the early 1960s. In 1961, artificial mRNAs were made and placed in a cell-free translation system. (Remember that the mRNA contains the information and instructions from the translation: *mRNA to Protein*). The first mRNA consisted only of uridine. The translation product was a polymer of phenylalanine (*phe*). Researchers concluded that the code contained within mRNA was a triplet code or Codon (a series of three adjacent bases in one polynucleotide chain). UUU encodes the amino acid *phe*. This can be re-stated as *UUU is the codon for phe. (Locate UUU in the chart below.)*
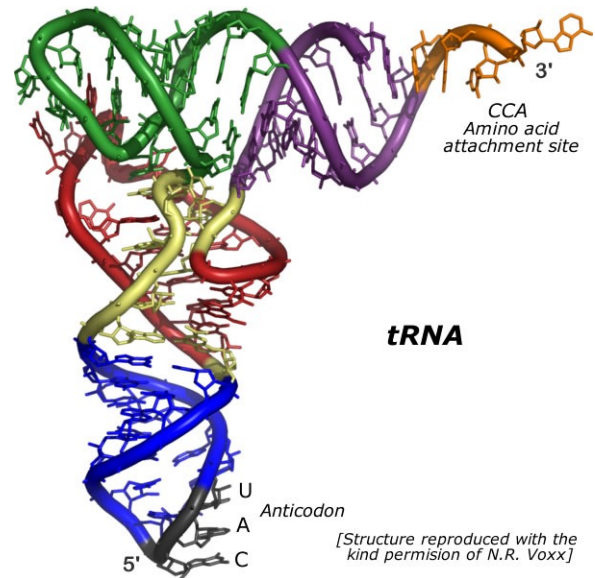

The Genetic Code   *Codon identification from mRNA*

The remainder of the code was rapidly deciphered. Creating all possible triplet codons from the 4 bases of RNA (U, C, A, and G) yields 64 codons, more than enough to encode the 20 common amino acids *(Phe, Leu, Ile, Met, Val, Ser…)*. AUG is the "start" codon and specifies methionine (met). *(Locate it on the chart and on the mRNA molecule.)* Three stop codons are UAA, UAG, and UGA. These are also known as nonsense codons as they do not code for any amino acid. Note in the chart that multiple codons specify the same amino acid. This makes the code redundant, but not ambiguous. Also, the genetic code is nearly universal.
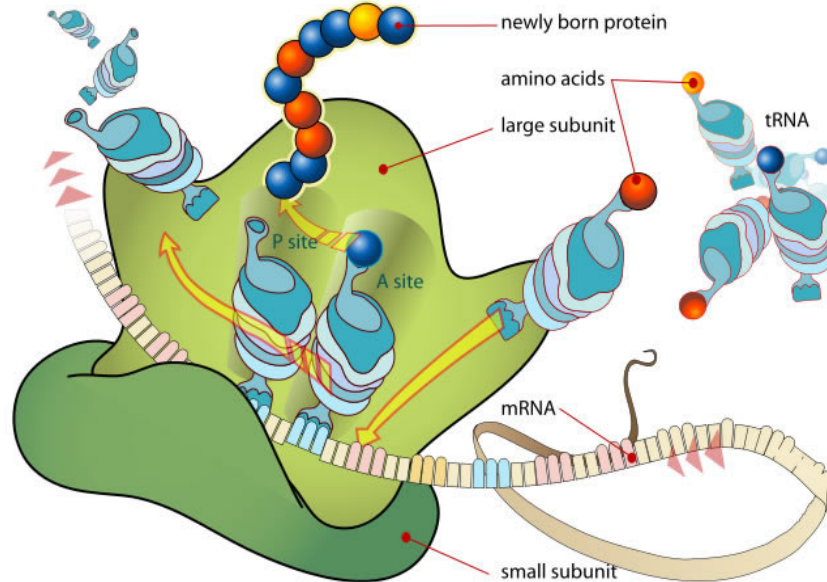
The Role of tRNA in Translation

From a previous discussion you learned that translation links the mRNA, appropriate tRNAs and ribosomes to produce the final product, a polypeptide. The tRNA *(see figure of tRNA)* contains an anticodon *(UAC)* and carries an amino acid *(CCA)*. The conformation of tRNA allows it to correctly interact with the ribosome.

It is of interest that the cell does not produce 61 different tRNAs. Researchers have found that the specificity for the base at the 3'-end of the codon (mRNA) and the 5'-end of the anticodon (tRNA) is not always strictly observed. This phenomenon is called "wobble". An example of wobble is the recognition of three alanine codons, GCA, GCC, and GCU, by the same tRNA.

tRNA

CCA
Amino acid attachment site

3'

U
A
C
Anticodon

5'

[Structure reproduced with the kind permision of N.R. Voxx]

Protein Translation (mRNA to Protein)


Protein Translation
Diagram showing the translation of mRNA and the synthesis of proteins by a ribosome
[Image courtesy of LadyofHats]

During protein translation amino acids are linked together to form a polypeptide chain which will later be folded into a protein. The ribosome is the workbench or factory for protein translation. It consists of a large and a small subunit *(see graphic – Protein Translation)*. A ribosome can use any mRNA and all species of tRNA to make a polypeptide product.

Translation begins with the formation of an initiation complex consisting of the small ribosomal subunit, mRNA and tRNAmet. Next, the large subunit binds and elongation begins. The anticodon of a tRNA molecule can form chemical bonds (i.e. base pair), with the mRNA's three letter codon. Thus the tRNA acts as the translator between mRNA and protein by bringing the specific amino acid coded for by the mRNA codon.[5] When the stop codon is reached, a protein release factor breaks the bond between the polypeptide and the tRNA on the ribosome.

Signal sequences often target the growing polypeptide to a specific cellular location. For example, a membrane protein is directed to the cell membrane as it would not want to be in the cytoplasm of the cell.

Most illustrations show a spaghetti-like strand of amino acids flowing away from a ribosome, but proteins begin folding as they are translated. Protein Folding is the process that converts a polypeptide into its characteristic and functional three-dimensional structure.
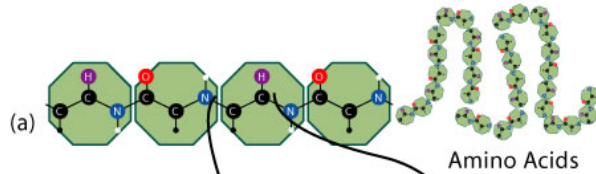

Protein Folding

Polypeptide chain
(output of Protein Translation)

3D structure
after folding
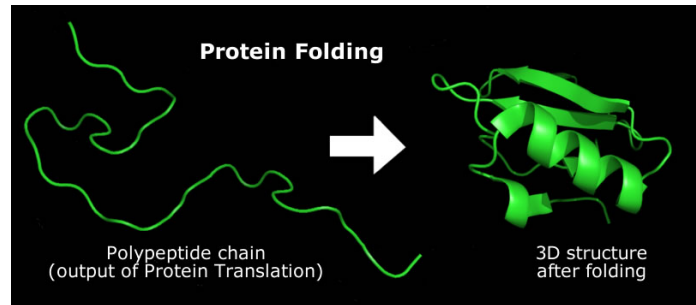
Protein Folding

The 4 levels of protein structure

Primary Structure
chain sequence
of amino acids

(a)

Amino Acids

Secondary Structure
occurs when the
sequence of amino
acids are linked by
hydrogen bonds

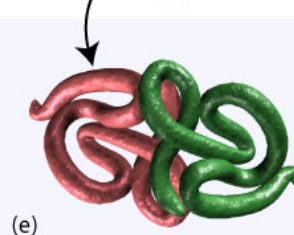(b)   Alpha Helix

(c)   Pleated Sheet

Tertiary Structure
occurs when certain
attractions are present
between alpha helices
and pleated sheets

(d)

Quarternary Structure
is a protein consisting
of more than one
amino acid

(e)

*The Four Levels of Protein Structure*
[Modified from freely available illustration provided by the Talking Glossary of Genetics]

Protein folding is complex, and can be described at four different levels *(refer to graphic above)*. The first level (Primary Structure) is simply the chain of amino acids (the output of protein translation). The second level (Secondary Structure) consists of the alpha-helix and the beta-pleated sheet. This level occurs when the sequence of amino acids are linked by hydrogen bonds. The third level (Tertiary Structure) consists of the additional folding and interactions between specific R-groups on amino acids, including disulfide bond formation, aggregation of hydrophobic side chains, van der Waals forces and ionic bond formation. The fourth level (Quarternary Structure) defines proteins that consist of multiple polypeptide subunits.

Proper folding is essential for protein function. If an enzyme misfolds, the active site may not form, and the enzyme would not be able to function as a biocatalyst. In similar ways, if a structural protein misfolds, it also would not be able to carry out a normal structural function.

BioMEMS, DNA and Proteins

The knowledge that we have gained about cells, DNA, RNA, proteins and the various biomolecules that make up human tissue has led to many new innovations. As previously mentioned, bioMEMS are used in the biomedical field for analyzing specific biomolecules in a sample (e.g., the home pregnancy test and insulin monitoring) and for delivering minute amount of drugs (e.g., insulin delivery). In addition, MEMS technology and nanotechnology are used to develop the biomedical assay (a device used to determine the amount of a particular constituent of a mixture, or of the potency of a drug). Such devices create surfaces that study proteins, DNA, various antibodies, and synthetic drug interactions with tissue. The following biochip was developed by Argonne National Laboratories. Each biochip has hundreds to thousands of gel drops on a glass, plastic or membrane substrate. The biochip system can identify infectious disease strains in less than 15 minutes when testing protein arrays and in less than two hours when testing nucleic acid arrays. [8]
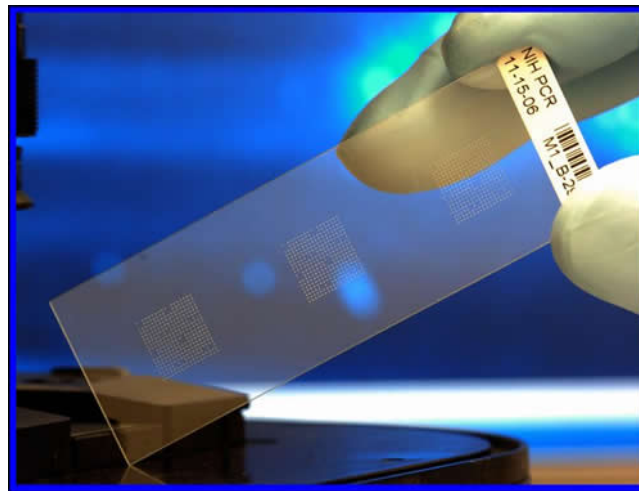
Biochip slide for testing protein arrays
[Image courtesy of Argonne National Laboratories]

MEMS technology and our knowledge of biomolecules have provided the ability to study protein-protein interactions on the basis of binding events. Such devices include peptide-loaded beads, microplates, pins and other flat micro-surfaces like membranes and chips. [9]

One such device is ELISA (Enzyme-Linked Immuno-Sorbent Assay). ELISA *(see figure)* is a sensitive immunoassay that uses an enzyme linked to an antibody or antigen as a marker for the detection of a specific protein, especially an antigen or antibody. [10] It has been widely used for detection and quantification of biological agents (mainly proteins and polypeptides). Its high selectivity and sensitivity draw great attractions in clinical, food safety, and environmental applications. [11]

BioLOC's CD-ELISA[TM]
[Printed with permission of BioLOC LLC]

Summary

- The Central Dogma of Biology states that genetic information flows from DNA to RNA to protein.
- Deciphering the genetic code has allowed scientists to translate the sequence of DNA into the sequence of amino acids that comprise the primary structure of a protein or polypeptide.
- Protein structure is related to function, and proteins fulfill a wide diversity of functions within cells.
- Proteins are also an integral part of many diagnostic aids and devices.
- Microarrays enable the identification and study of specific proteins within small samples.

Food For Thought / Answers

1. How does RNA differ from DNA?

2. DNA is composed of 2 strands, of which only one strand is used as a template for RNA synthesis for a specific gene. By what mechanism is the correct strand chosen?

3. What is a codon?

4. How are proteins directed to the correct compartment within a cell?

5. Given the types of chemical interactions that promote and maintain protein structure, can you think of ways to disrupt or break those interactions?

References

[1] Fundamentals of BioMEMs and Medical Microdevices. Steven S. Saliterman. Wiley Interscience. 2006.

[2] Life: The Science of Biology, 8th edition, Sadava et.al. Freeman. 2007.

[3] Bioinquiry, 3rd edition, Pruitt and Underwood. Wiley. 2006.

[4] Genetics: From Genes to Genomes, Hartwell et.al., McGraw Hill. 2008.

[5] Translation. Nobleprize.org. http://bit.ly/2xwAzez

[6] "Learn Genetics". University of Utah. http://learn.genetics.utah.edu

[7] The BioTechnology Project at MATC. Madison Area Technical College.

[8] "Biochip technology could become standard diagnostic tool for human, veterinary medicine". Argonne News. Dec 20, 2006.

[9] "Synthetic peptides, peptide pools and peptide arrays for the analysis of B- and T-cell responses". Holger Wenschuh, PhD. JPT Peptide Technologies, Berlin.

[10] "ELISA". The Free Dictionary. http://medical-dictionary.thefreedictionary.com/ELISA

[11] "CD-ELISA". BioMEMS/NMEMS. Lai, Siyi. 2004 http://www.sciencedirect.com/science/article/pii/S0065242306420072

Related SCME Learning Module, Activity and Assessment

- DNA Overview Learning Module
- DNA to Protein Activity
- DNA to Protein Assessment

Glossary of Key Terms

Alkaptonuria - An inherited condition that causes urine to turn black when exposed to air. Ochronosis, a buildup of dark pigment in connective tissues such as cartilage and skin, is also characteristic of the disorder.

Amino acid - The subunit, building blocks of proteins. There are 20 commonly occurring amino acids.

Antibody - A specific subset of proteins made in response to an antigen (a foreign substance) as part of the immune response.

Anticodon - Nucleotide triplets on tRNA that recognize codons on the mRNA by complementary base pairing and wobble.

ChIP/chip analysis - Also known as genome wide localization, this technique identifies all the genomic sites at which a transcription factor expressed in a particular cell type may bind to DNA.

Codon - A nucleotide triplet that specifies a particular amino acid to be inserted in a growing polypeptide chain during translation.

Enzyme - A molecule that functions as a biological catalyst, and lowers the energy of activation of a metabolic reaction. Most, but not all, are protein in nature, and activity depends on their maintaining a specific conformation.

Ferment - Any of a group of living organisms, as yeasts, molds, and certain bacteria, that cause fermentation.

Genetic code - The information contained in a gene within the DNA that specifies the synthesis of a protein. The code refers to the sequence of bases that ultimately determine the amino acid sequence of a protein.

Promoter - A region of DNA sequence to which RNA polymerase or associated transcription factors bind.

Protein - Large polymer composed of hundreds to thousands of amino acid subunits linked together in a specific order into long chains. Proteins are required for the structure, function, and regulation of an organism's cells, tissues and organs.

Proteome - The complete set of proteins encoded by a genome.

Proteomics - A complete analysis of all, or most, of the proteins in a particular cell type or organism.

Ribonucleotide - The subunit building block of RNA.

Ribosome - Cytoplasmic structures composed of ribosomal RNA (rRNA) and proteins. They are sites of protein synthesis.

RNA Polymerase – An enzyme that produces RNA. A polymerase that catalyzes the synthesis of a complementary strand of RNA from a DNA template, or, in some viruses, from an RNA template.

Template - A strand of DNA or ENA that is used as a model by DNA or RNA polymerase or by reverse transcriptase for the creation of a new complementary strand of DNA or RNA.

Translation - The process in which the mRNA directs the synthesis of a polypeptide from amino acids according to the genetic code.

Transcription - The conversion of DNA-encoded information to an RNA equivalent.

Transcription factor - A protein or RNA whose binding or indirect association with a promoter helps regulate the timing, location and level of a specific gene's transcription.

Wobble - The ability of the 5'-most nucleotide of an anticodon to interact with more than one nucleotide at the 3'-end of codons. This helps explain the degeneracy of the genetic code.

Disclaimer

The information contained herein is considered to be true and accurate; however the Southwest Center for Microsystems Education (SCME) makes no guarantees concerning the authenticity of any statement. SCME accepts no liability for the content of this unit, or for the consequences of any actions taken on the basis of the information provided.