

Background

Proteomics, Evolution, and Classification of Fish

Proteomics

Proteomics is the study of proteins, particularly their structures and functions. This term was coined to make an analogy with genomics, and while it is often viewed as the "next step", proteomics is much more complicated than genomics. Most importantly, while the genome is a rather constant entity, the proteome differs from cell to cell and is constantly changing through its biochemical interactions with the genome and the environment. One organism will have radically different protein expression in different parts of its body, in different stages of its life cycle, and in different environmental conditions.

The entirety of proteins in existence in an organism throughout its life cycle, or on a smaller scale the entirety of proteins found in a particular cell type under a particular type of stimulation, are referred to as the *proteome* of the organism or cell type, respectively.

With completion of a rough draft of the human genome, many researchers are now looking at how genes and proteins interact to form other proteins. The large increase in protein diversity is thought to be due to alternative splicing and posttranslational modification of proteins. This discrepancy implies that protein diversity cannot be fully characterized by gene expression analysis alone, making proteomics a useful tool for characterizing cells, tissues, and organisms of interest.

To catalog all human proteins and ascertain their functions and interactions presents a daunting challenge for scientists. An international collaboration to achieve these goals is being coordinated by the Human Proteome Organization (HUPO).

Evolution

The term evolution probably brings to mind Charles Darwin and the Theory of Natural Selection. In short, this theory states that there are more organisms brought into the environment than can be supported by the environment. Each of these individuals are different – even among the same species. The environment selects organisms best suited to survive and reproduce based on those differences. Adaptations are the differences that make one organism more suited to the environment than another individual. These adaptations are phenotypic (physical) characteristics such as finch beaks that are determined by a genetic component. The genetic component is inherited from the parent in the form of genes.

Variations in an organism's proteins may reflect physiological adaptations to an ecological niche and environment, but they originate as chance DNA mutations. Such random mutation events, if favorable, persist through the natural selection process and contribute to the evolution of new species – with new specialized functions.

The discovery of the chemical structure of DNA by Watson, Crick, Wilkins, and Franklin and our understanding of how the triplet code of nitrogen bases leads to the synthesis of proteins (which is the phenotypic expression) convinced us that adaptations are the result of changes in the DNA code (mutations). However, current research in the field of proteomics is leading some scientists to question whether or not DNA is the final determining factor in the synthesis of proteins and thus the determining factor in evolution.

The central dogma of molecular biology of DNA→RNA→protein has given us a comfortable explanation of how the information encoded by our DNA is translated and used to make an organism. It describes how a gene made of DNA is transcribed by messenger RNA and

then translated into a protein by transfer RNA in a complex series of events utilizing ribosomal RNA and amino acids. New discoveries about alternative roles for RNA, multiple forms of proteins being encoded by single genes in our cells, and changes to proteins after translation are changing this comfortable scenario and we are finding that things (as ever in biology) are not so simple. Although in essence the central dogma remains true, investigations into genomics and proteomics are revealing a complexity that we had never imagined.

In 1990, a massive research effort took place to sequence what was estimated to be the 100,000 genes that coded for each protein synthesized by humans (the human genome). This study, the Human Genome Project, took 13 years to complete. When the study began, scientists estimated that there were over 100,000 human genes. Now, years after the genome has been sequenced, there is still no consensus on the actual number of human genes, but the current estimate is down to around 22,000 human genes, this only a few thousand more genes than encodes the genome of a much simpler organism, *C. elegans*, a nematode worm that has around 19,000 genes.

So why are a similar number of genes required to make a worm and a person? Importantly, a human has a much larger total genome (3 billion base pairs) than a worm (100 million base pairs) suggesting that the total amount of DNA rather than the actual number of genes may be what gives rise to complexity. In addition, recent developments have shown it is quite common in complex organisms for a single gene to encode multiple proteins. Moreover, changing when, to what level and where a protein is expressed, or changing a protein after it has been translated (posttranslational modification) can result in proteins with very different functions. This realization of the importance and diversity of proteins started a whole new field termed **proteomics**.

Proteomics was initially defined as the effort to catalog all the proteins expressed in all cells at all stages of development. That definition has now been expanded to include the study of protein functions, protein-protein interactions, cellular locations, expression levels, and posttranslational modifications of all proteins within all cells and tissues at all stages of development. Thus, it is hypothesized that a large amount of the noncoding DNA in the human genome functions to highly regulate protein production, expression levels, posttranslational modification etc., and it is this regulation of our complex proteomes, rather than our genes, that makes us different from worms.

Research in the proteomic field has discovered a number of modification systems that allow one gene to code for many proteins and mechanisms that finely regulate the sub- and extracellular locations and expression levels of proteins. These include alternative splicing of exons, use of different promoters, posttranscriptional modification, translational frameshifting, posttranslational modification, and RNA editing.

Muscle Proteins

Our most familiar daily movements, from walking to simply breathing, are driven by the interactions between specialized proteins in our muscle fibers. The basic contractile elements of the muscle cells are the myofibrils that are bundled into muscle fibers. Each myofibril consists of a linear series of contractile units called sarcomeres.

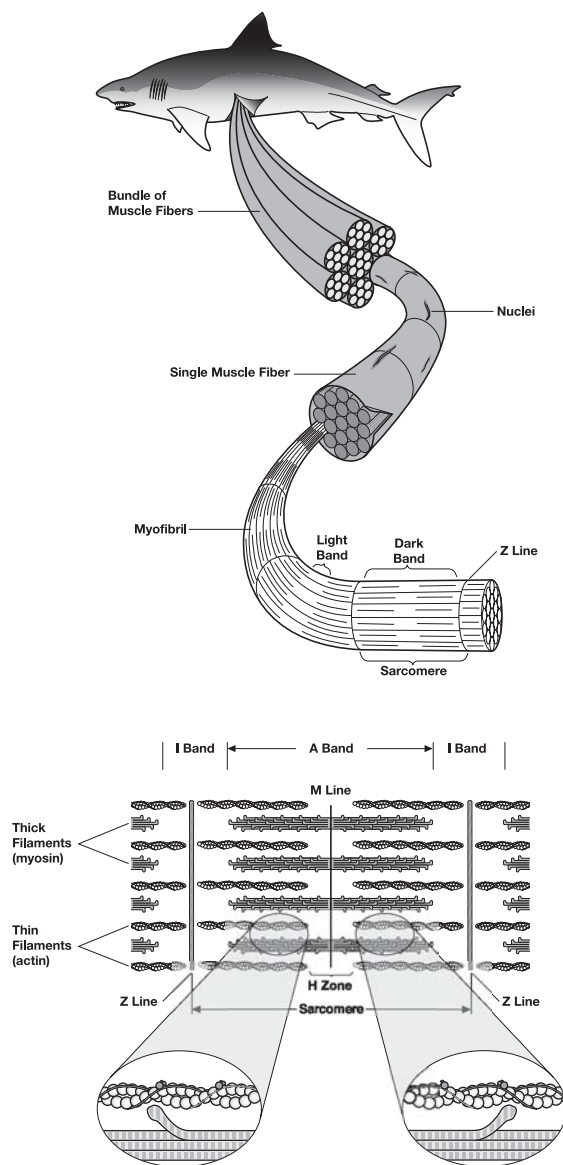


Fig. 12. Telescopic view of muscle structure: Thick myosin filaments and thin actin filaments form myofibrils, which are bundled together to make muscle fibers. (Figure modified from Campbell 1996 with permission.)

Sarcomeres are precisely arranged assemblies of actin and myosin protein filaments. Thin filaments of actin are aligned with thick filaments of myosin in a parallel and partly overlapping manner. The sarcomere shortens when myosin hydrolyzes ATP to slide along the actin filament, pulling the ends of the sarcomere towards each other. The combined contraction of many sarcomeres along a muscle fiber causes contraction of the entire muscle. It is important to note that, although actin and myosin are the major components, other proteins are also found in muscle tissue.

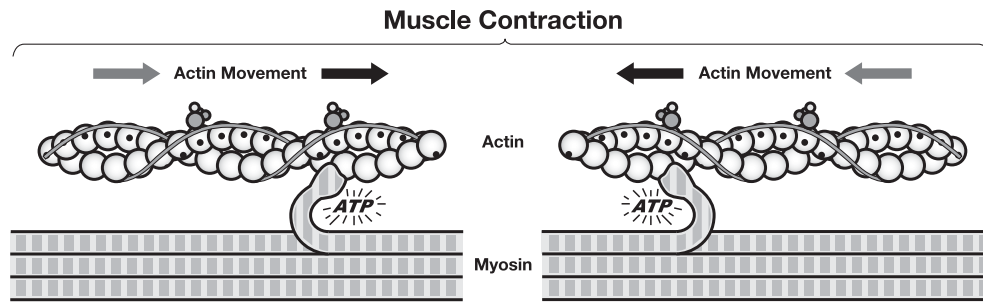


Fig. 13. Hydrolysis of ATP causes myosin and actin filaments to slide past one another, shortening the sarcomere and contracting the muscle. (Figure modified from Campbell 1996 with permission.)

Other Muscle Proteins

Numerous proteins besides actin and myosin are also required for muscle contraction (please refer to the table below). While actin and myosin are highly conserved across all animal species, other muscle proteins show more variability. These variations in an organism's muscle proteins may reflect refinements of muscle function and performance that are adaptive to particular niches, environments, or physiological stresses.

Table 2. Characterized muscle proteins, in order of decreasing size, adapted from Alberts et al. (1994).

Protein	MW (in kD)	Function
titin	3,000	centers myosin in sarcomere
dystrophin	400	anchoring to plasma membrane
filamin	270	crosslinks actin filaments into a gel
myosin heavy chain	210	slides actin filaments
spectrin	265	attaches filaments to plasma membrane
M1/M2	190	myosin breakdown product
M3	150	myosin breakdown product
C protein	140	myosin breakdown product
nebulin	107	regulates actin assembly
α -actinin	100	bundles actin filaments
gelsolin	90	fragments actin filaments
fimbrin	68	bundles actin filaments
actin	42	forms filaments
tropomyosin	35	strengthens actin filaments
troponin (T)	30	regulates contraction
myosin light chains	15–25	slide actin filaments
troponin (I)	19	regulates contraction
troponin (C)	17	regulates contraction
thymosin	5	sequesters actin monomers

Lesson 1: Introduction to Protein Electrophoresis and SDS-PAGE

How Can We Study Proteins Found in Muscle?

Polyacrylamide gel electrophoresis (PAGE) can be used to separate small molecules such as proteins. Understanding protein structure is important to understanding how we can use PAGE for protein analysis.

Proteins are made of smaller units (monomers) called amino acids. There are 20 common amino acids. The sequence and interaction between these different amino acids determine the function of the protein they form. Amino acids are joined together by peptide bonds to form polypeptide chains. Chains of amino acids constitute a protein. In turn these chains may interact with other polypeptides to form multi-subunit proteins.

Amino acids can be combined in many different sequences. The sequence of the amino acids in the chain is referred to as the primary protein structure. All amino acids have the same basic structural component (Figure 15).

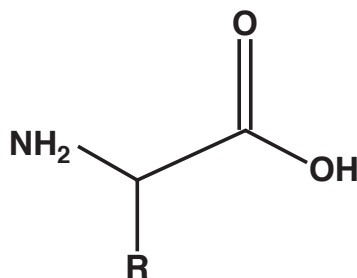


Fig. 15. Chemical structure of an amino acid.

The "R" group may be charged or uncharged, or may be a long side chain. Thus, each amino acid has different properties and can interact with other amino acids in the chain. Hydrogen bonding between these side chains, primarily between the C=O and the N-H groups, causes the protein to bend and fold to form helices, pleated sheets, reverse turns, and non-ordered arrangements. Disulfide bonds between methionines can also bend and loop the amino acid chain. This is considered the secondary structure of the protein.

The tertiary structure of the protein is determined by the interaction of the hydrophilic and hydrophobic side chains with the aqueous environment. The hydrophobic regions aggregate to the center of the molecule. The hydrophilic regions orient themselves toward the exterior. These ordered bends and folds make the protein compact. Examples of tertiary protein structure are structural and globular proteins.

The quaternary structure of proteins is achieved from the interaction of polypeptide chains with others. Multiple polypeptides can combine to form complex structures such as the muscle protein, myosin, or the blood protein, hemoglobin, which are both composed of four polypeptide chains. These complex proteins are often held together by disulfide bonds between methionines. In fact, PAGE analysis was first carried out in 1956 to show the genetic disease sickle cell anemia is caused by a change to a single amino acid of the hemoglobin protein (Ingram 1956).

Prior to electrophoresis, the proteins are treated with the detergent sodium dodecyl sulfate (SDS) and heated. SDS and heat denatures (destroys) the protein tertiary and quaternary structures, so that the proteins become less three dimensional and more linear. SDS also gives the protein an overall negative charge with a strength that is relative to the length of its polypeptide chain, allowing the mixture of proteins to be separated according to size.

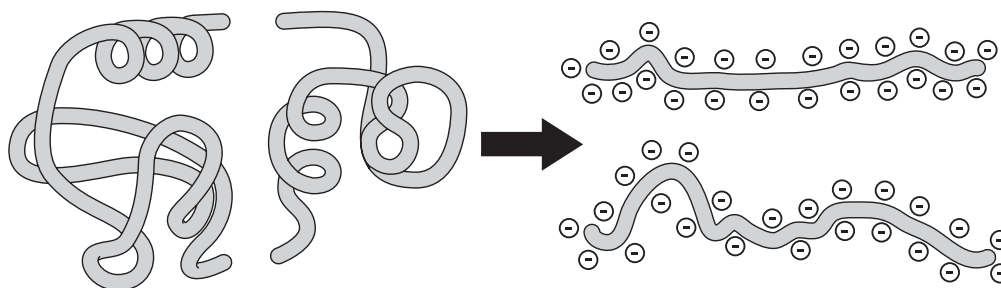


Fig. 16. The combination of heat and the detergent SDS denatures proteins for SDS-PAGE analysis.

The proteins, in their SDS-containing Laemmli sample buffer, are separated on a gel with a matrix that acts to sieve the proteins by size upon addition of an electric current. A polyacrylamide gel is positioned in a buffer-filled chamber between two electrodes, protein samples are placed in wells at the top of the gel, and the electrodes are connected to a power supply that generates a voltage gradient across the gel. The SDS-coated, negatively charged proteins migrate through the gel away from the negatively charged anode toward the cathode, with the larger proteins moving more slowly than the smaller proteins. This technique was developed by U.K. Laemmli, whose 1970 Nature paper has received the highest number of citations of any scientific paper. SDS-PAGE is still the predominant method used in vertical gel electrophoresis of proteins.

As soon as the electric current is applied, the SDS-coated proteins begin their race toward the positive electrode. The smaller proteins can move through the gel more quickly than the larger ones, so over time, the proteins will be separated according to their sizes.

Protein size is measured in **daltons**, a measure of molecular mass. One dalton is defined as the mass of a hydrogen atom, which is 1.66×10^{-24} gram. Most proteins have masses on the order of thousands of daltons, so the term **kilodalton (kD)** is often used to describe protein molecular weight. Given that the average weight of an amino acid is 110 daltons, the number of amino acids in a protein can be approximated from its molecular weight.

- Average amino acid = 110 daltons
- Approximate molecular weight of protein = number of amino acids x 110 daltons

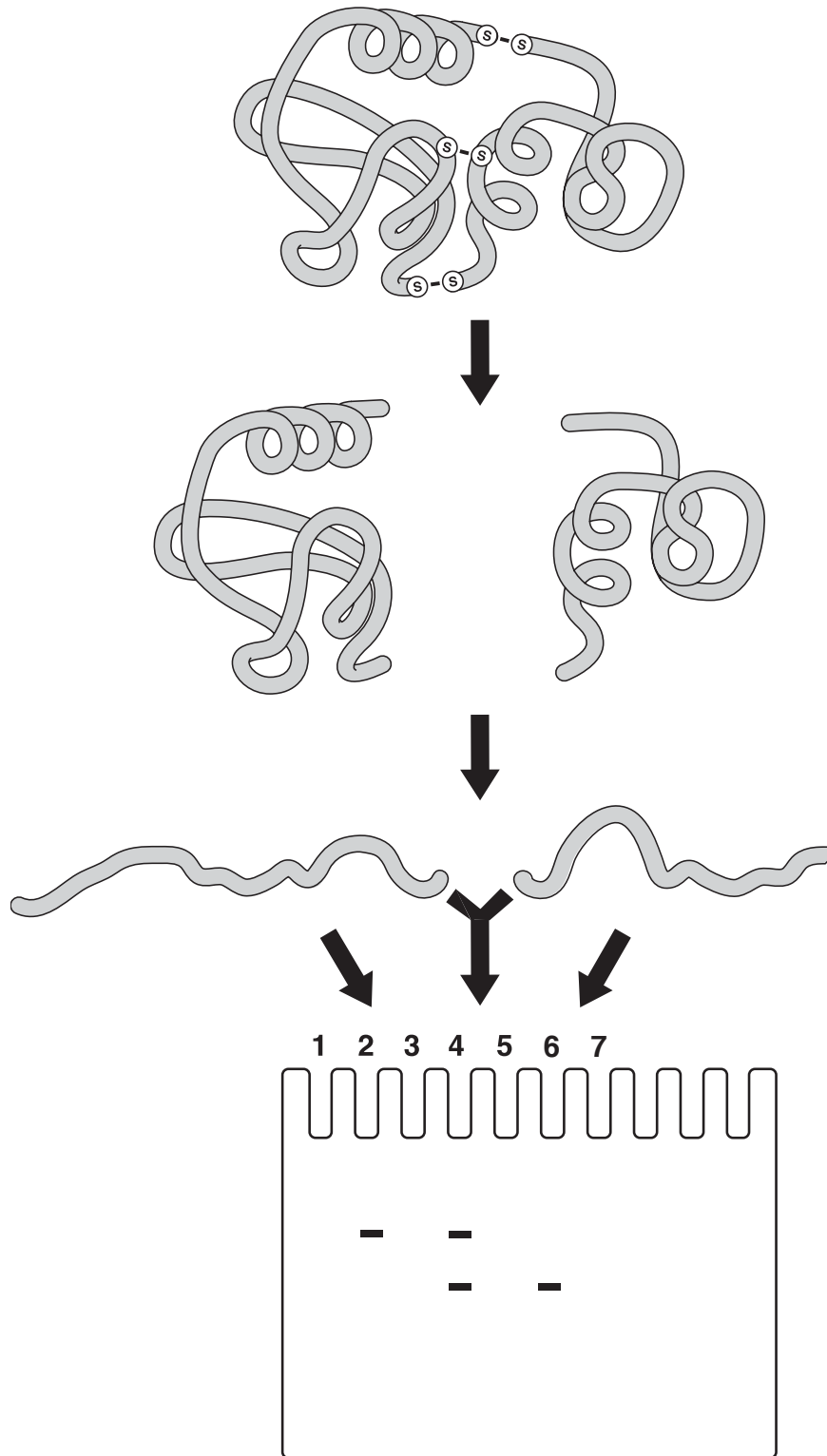


Fig. 17. A quaternary protein complex denatured with reducing agents, heat, and SDS, can be separated into individual proteins and resolved by size using SDS-PAGE.

In this investigation, you will use SDS-PAGE to separate and analyze the protein profiles of the muscle tissue of different fish. By comparing the protein profiles of different fish species you can test the hypothesis that protein profiles are indicators of genetic and evolutionary relatedness.

Visualizing your proteins

Proteins in your samples are not visible while the gel is running. The only visible proteins will be those in the Precision Plus Protein Kaleidoscope standard that have been prestained with covalently attached dyes. You should see these proteins resolve into multicolored bands that move down the gel as power is run through the gel. If the electric current is left on for too long, the proteins will run off the bottom of the gel. To guard against this and to show you the progress of your gel if you did not have the standards, a blue tracking dye is mixed with the Laemmli sample buffer used to prepare your protein samples. This blue dye is negatively charged and is also drawn toward the positive electrode. Since the dye molecules are smaller than the proteins expected in most samples, they move ahead of the proteins in the gel.

After turning off the electric current and removing the gel from the glass plates that hold it in place, the gel is placed in a stain. The stain used in this technique was originally developed to dye wool, which, like your own hair, is composed of protein. This stain binds specifically to proteins and not to other macromolecules such as DNA or lipids. After destaining, distinct blue bands appear on the gel, each band representing on the order of 10^{12} molecules of a particular protein that have migrated to that position: the larger the amount of protein, the more intense the blue staining.