A Is For . . .

Proteome

proteome All the proteins expressed by a genome, generally, or all the functioning proteins present in a cell, particularly. Proteomics is the study of the proteome, as genomics studies the genome.

Sequencing the human genome is not yet "finished" and may never be, but regardless of how many genes there are or which ones are specific to humans, it is clear that one gene makes more than one protein. In humans, the ratio is at least 3 to 1, and it may be 20 to 1. It is these functioning gene products rather than genes themselves that need to be examined if the operation of a cell and its disorders in disease are to be comprehended.

Making a "protein inventory" of a cell may be simple in concept, but it is a long way from happening. Proteins often function not as separate entities but in complexes with other proteins. An amino acid sequence gives only imperfect clues to a protein's folded conformation, which is one of the main characteristics determining the nature of each protein's interactions with others.

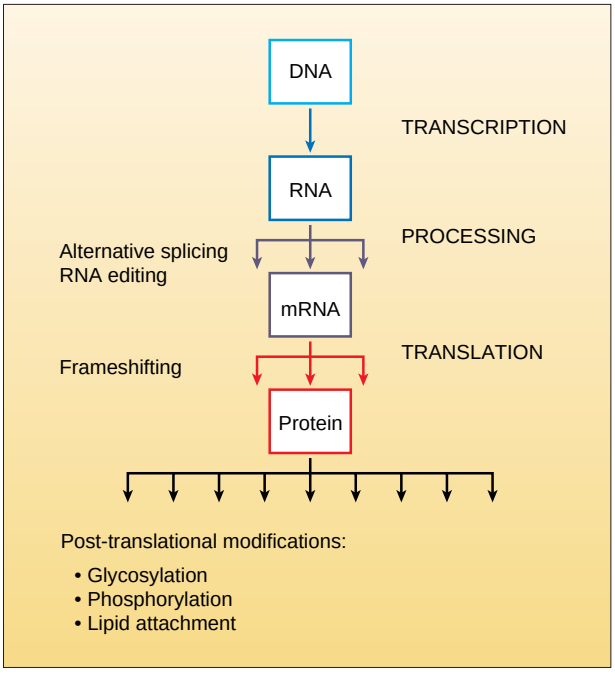
Many kinds of modifications are known to affect proteins after they have been made by translation from messenger RNA. Perhaps the most common is glycosylation, which in the case of some proteins can take place at a number of different sites, in effect giving rise to functionally different molecules.

Procedures for identifying proteins within a tissue or serum sample have become increasingly sophisticated and sensitive. The process still begins with two-dimensional gel electrophoresis, which separates proteins according to their charge (the first dimension) and molecular mass (the second dimension).

As many as 3000 different proteins can be separated on each gel, but there are many more than that in a single cell at any one time. Which ones are found depends on how the sample is obtained and on the pH range of the gel strip. It is possible, for example, to start with proteins that are already known to share a certain biochemical characteristic such as hydrophobicity, or those that are localized to specific cell compartments

To discover what protein constitutes a particular spot on a gel, the spot is cut out of the gel and digested with trypsin into a mixture of peptides. These are ionized so that their masses can be measured by mass spectrometry. The result is compared by computer software with peptide masses predicted from theoretical digestion of known protein sequences. Failure to find a match requires discovery of enough actual amino acid sequence (by tandem mass spectrometry) to allow the design of nucleotide probes that can identify new proteins.

As genes are sequenced, theoretical functions are assigned to their protein products by computer-based similarity searches against proteins of known function, many of which were originally characterized in bacteria. In organisms with sequenced genomes, no related proteins can be identified for about one third of the sequences; the percentage will undoubtedly be higher in the human genome.



A single gene can make a number of proteins in a variety of ways. By far the most numerous are post-translational modifications, more than 200 kinds of which have been described. The functional state of a protein varies with these modifications and may also vary over time.

All such proteins will need to be expressed and purified in vitro so that their three-dimensional structures can be predicted by X-ray crystallography and nuclear magnetic resonance. Those structures usually offer clues to possible interactions with other proteins and may reveal catalytic sites.

Meanwhile, samples taken from normal and diseased tissue can be compared, an approach similar to "microarray" techniques already used to detect messenger RNA. The presence or absence of a particular protein, or its presence in different forms, can suggest possible roles for genes in disease.

Proteins present different technical problems than nucleic acids. There is no protein amplification technique comparable to the polymerase chain reaction, and because protein properties result from their folded structures, analytic methods are difficult to generalize. Most importantly, the variety of post-translational modifications remains of largely unknown significance.

RECENT REVIEW

Rosamonde E. Banks et al.: Proteomics: New perspectives, new biomedical opportunities. *Lancet* 356:1749-1756, November 18, 2000.