Research Methods

14 Molecular Biology

Francesco Emma · Luisa Murer · Gian Marco Ghiggeri

Introduction

The field of molecular biology developed initially between the 25 years from 1940–1965, through the efforts of scientists which led to the discovery of the genetic code and the establishment of the role of RNA in the synthesis of proteins (1, 2). Thereafter, efforts of an increasing number of researchers have provided an explosion of key technological innovations allowing recombinant DNA technology to be applied to a wide variety of biologic problems and to enter clinical practice. Major milestones include discovery of restriction nucleases and DNA ligases, development of DNA libraries, DNA cloning procedures, nucleic acid hybridization techniques, the polymerase chain reaction (PCR), rapid sequencing techniques, and the production of transgenic animals. These advances permitted detection, amplification and engineering of DNA sequences, and delineation of the role of genes in cellular physiology and pathophysiology.

This molecular revolution has affected all sciences, including medical and clinical research, and has culminated in the completion of the human genome project (3). It has provided new insight into the complexity of living organisms that are now been studied by functional genomic and proteomic approaches that involve microengineering, computer technology, and bioinformatics.

This chapter illustrates the basic concepts of molecular biology that underlie various strategies that have been developed for studying living cells and organisms and summarizes molecular biology techniques that are used in nephrology research and clinical practice. It is therefore necessarily limited in breadth and depth. For more information, the reader can refer to the many specialized reference textbooks that are available (1, 2).

Basic Concepts

From DNA to Proteins

In the mid-1950s, the "central dogma of molecular biology" was proposed soon after the discovery of the DNA structure in 1953. To date, this dogma remains one of the keystones that guides the study of genetic human diseases.

DNA sequences are transcribed into RNA molecules that carry the flow of genetic information out of the nucleus to be translated into proteins. These processes have been well defined and are detailed in most introductory cell biology texts (1, 2). This simple concept has been however thoroughly revised, as genetic information has been shown to be conveyed in both directions through a complex series of feedback loops.

The link between nucleic acids and proteins is contained in the genetic code. Sixty-one of the 64 codons correspond to amino acids, whereas three correspond to termination codons. The code is said to be "degenerate", as it contains redundancies. Codons corresponding to the same amino acid generally differ by the nucleotide in the third position.

As a practical consequence of the degeneracy of the genetic code, nucleotide substitutions (point mutations) in the third nucleotide of a given codon may not change the primary sequence of a protein and are frequently found in nature, as they are not subject to natural selection. Point mutations in the first and second position result in an amino acid substitution or in a termination codon, either of which can dramatically alter protein structure and function.

In principle, each messenger RNA (mRNA) can be translated in any of the three possible reading frames determined by overlapping triplet codes. With few exceptions, only one reading frame produces a functional protein because stop codons encountered in the other two reading frames terminate translation. As the only punctuation signal is located at the start codon (ATG), the reading frame is set at the beginning of the translation and proceeds until a termination codon is reached. Thus, finding the correct reading frame and locating the start codon are essential steps in the process of cloning genes and defining their protein products. Mutations causing deletion or insertion of one or two nucleotides result in a shift of the reading frame (frameshift mutations) and cause production of aberrant protein products.

© Springer-Verlag Berlin Heidelberg 2009

Gene Structure

Complementary strands of chromosomal DNA are arranged in an anti-parallel fashion as dictated by hydrogen bond pairing of nucleotide bases. It is estimated that the human genome is composed of approximately 3 billion base pairs of DNA containing over 30,000 genes, arrayed on 23 pairs of chromosomes (4). Overall, the amount of DNA containing genes comprises a minority of nuclear DNA.

Traditionally, a gene is depicted as a "transcriptional unit" as illustrated in *Fig. 14-1*. The DNA double helix is represented as a line interrupted by rectangular boxes corresponding to exons, with its 5′ end on the left and its 3′ end on the right. Each gene is divided into two major regions, namely the promoter and the coding regions. The promoter region is located upstream near the 5′ boundary of the coding region and contains clusters of short sequences (less than ten base pairs) spread over a few hundred bases that bind transcription factors. These regulatory proteins mediate the attachment and activation of type II RNA polymerase, which mediates transcription along the DNA while unwinding the duplex and adding nucleotides to the growing RNA molecule. DNA sequences in the promoter region that bind to transcription factors are referred as *cis*-acting elements. In some

Figure 14-1

General organization and processing of a eukaryotic gene. A gene consisting of four exons and three introns is shown. The promoter region (black box) is located near the 5′ end of the gene. Transcription of the gene yields a primary mRNA transcript that contains both exon and intron sequences. Differential splicing of this mRNA transcript yields two mature mRNA species by inclusion or exclusion of exon 3 (hashed box). In order to be exported into the cytosol and translated, mRNAs need to be modified by polyadenylation of their 3′ end and capping of their 5′ end with methylated guanosines. Differential splicing results in two different protein isoforms which are encoded by a single gene.

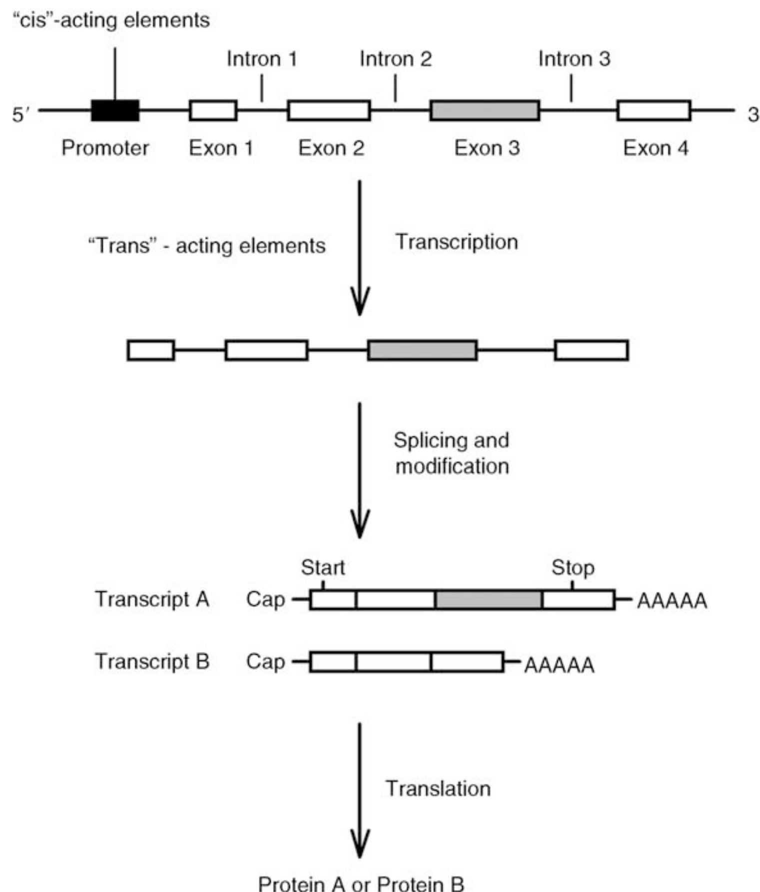
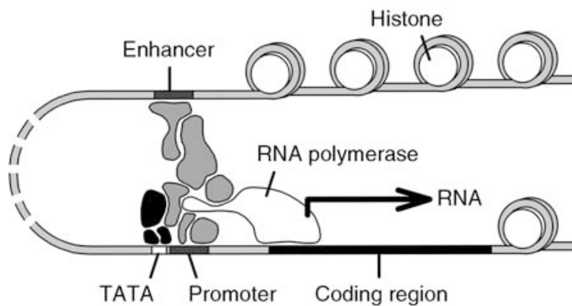


Figure 14-2

Transcriptional regulation in eukaryotes. Different transcription factors bind to the promoter and enhancer regions and position type II RNA polymerase at the starting point of transcription. General transcription factors shown in black interact with the TATA box and hold together the enhancer-promoter-RNA polymerase complex. When not activated, eukaryotic genes are hidden in nucleosomes which are composed of a central core formed by histones. Activation of eukaryotic genes requires remodeling of the chromatin in order to allow the transcriptional apparatus to interact with regulatory sequences.



cases, transcription factors also interact with other *cis*-acting elements (enhancers) that are located at a greater distance (up to a few thousand bases) from the promoter. The physical gap between the enhancer and the promoter explains the need for accessory factors that convey transcriptional signals to the RNA-polymerase through protein-protein interactions (*Fig. 14-2*). All molecules, generally proteins, which are physically involved in the regulation of transcription are referred to as *trans*-acting elements, because their DNA sequence resides in a different location of the genome (1, 2). The coding region contains information for protein synthesis. In this region, most genes are composed of a succession of exons and introns. Exons are coding sequences that ultimately transfer into mature mRNA, whereas introns are edited out of the newly synthesized transcript by a process called splicing.

Control of Gene Expression

As only a fraction of the available genes are expressed in a given cell at a given developmental stage, differential transcription and processing of genes provides for enormous diversity between cells within the same organism.

Although gene expression can be controlled at different levels, transcriptional regulation generally predominates. Information that governs transcription is located in DNA sequences that correspond to *cis*-acting elements or which encode for transcription factors. These sequences occupy a minimal portion of the entire genome but are key determinants of cell organization by ensuring a balanced expression of different genes that preserves phenotypic stability. They also provide for differences among cell types within the same organisms and are at the core of the evolutionary process.

Gene expression in eukaryote organisms is complex. Unlike prokaryotes, genes operating within the same metabolic pathway are not usually physically aligned along the genomic DNA and are often located on different chromosomes. The expression of functionally related genes is achieved by activation of shared classes of transcription factors that act on common structural motifs in DNA-binding domains (enhancers and promoters). Transcription factors are also often involved in developmental regulation. The eukaryotic type II RNA polymerase differs from its prokaryotic ancestor by the complexity of transcription factors that are required for its activation (5) (*Fig. 14-2*). This elaborate modular system allows for flexible and highly coordinated regulation of gene transcription. In transcription, DNA rearranges to allow interaction between transcription factors located at the promoter and at the enhancer site. The TATA box, which contains consensus sequences recognized by general transcription factors (TFIID, TBP, and TAF), is usually located approximately 24 bases upstream from the promoter and acts as a bridge that holds together the enhancer-promoter-RNA polymerase complex and positions the enzyme at the starting point of transcription (6). Although enhancers are not essential, they increase promoter efficiency.

A second characteristic of transcription in eukaryotic cells is related to the association of nuclear DNA with histones, forming nucleosomes. This particular organization prevents interaction between regulatory sequences and transcription factors unless conformational changes of the chromatin permit gene activation (*Fig. 14-2*).

Gene transcription can be chemically blocked in eukaryotic cells by methylation of specific DNA regions located near the promoter. Experimentally, the state of DNA methylation can be determined by inhibition of the nuclease activity of restriction enzymes that recognize non-methylated GC doublets.

Transcription Factors

The major characteristic of transcription factors is to contain specialized domains that allow for interaction with the DNA double helix.

Zinc fingers, helix-turn-helix (HTH) domains, helix-loop-helix domains, and leucine zippers are examples of DNA binding domains that are encountered in more than 80% of transcriptional factors (7). Zinc finger motifs are composed of a Zn ion holding together a peptide loop (the finger) through interaction with two histidines and two or four cysteines. The functional DNA binding domain is located in the amino acid residues at the base of the finger and recognizes specific consensus sequences that are generally located in the enhancer region. Steroid receptors are the most popular examples of this class of transcription factor, including glucocorticoid, mineralocorticoid, androgen, progesterone, thyroid hormone, and vitamin D receptors (8).

Similar DNA binding activity characterizes other classes of transcription factors. HTH domains are often found in proteins that have key roles during morphogenesis such as the homeobox group of transcription factors (9). Members of the helix-loop-helix class include proteins that control myogenesis, such as MyoD, that have been implicated in the transdifferentiation of myofibroblasts, which promote renal fibrosis (10). Leucine-zipper motifs are used by the Jun and Fos proteins. Both are members of the AP-1 heteromeric transcription complex that is involved in cell proliferation and regulation of cell matrix during fibrogenesis (11). An important transcription factor that is often implicated in renal diseases is nuclear factor-κB, which regulates many proinflammatory pathways and is itself under the control of other proteins, such as angiotensin II, known to promote inflammatory reactions in the renal parenchyma (12).

Transcriptional Control during Development

Transcriptional regulation plays a central role during development, by governing complex sequences of events that transform undifferentiated embryonic cells into highly specialized mature cells (1, 2). Three major classes of developmental regulatory genes have been identified. These include maternal, segmentation, and homeotic (Hox) genes. Each are expressed at different stages of cell maturation following complex sequences of activation in which gene products that are expressed at a given stage activate genes at a following stage. This highly sophisticated sequence of events creates a hierarchy according to which maternal, segmentation, and homeotic genes are expressed sequentially. Maternal systems activate expression of specific transcription factors, termed morphogens, that are responsible for the initial patterning of the embryo. Segmentation genes are zinc-finger proteins that control the boundaries of compartments, whereas Hox genes control the differentiation steps of each segment. Hox genes are characterized by a common HTH-type DNA-binding motif at their carboxy- terminus termed homeobox and are organized in the human genome in four different clusters (A,B,C,D) containing up to ten genes each. As the number of identified developmental transcription factors is rapidly increasing, their impact on renal development and involvement in congenital renal anomalies is a field of active research (13). Moreover, during the recovery phase of renal cell injury, cells reactivate several developmental genes in repair and restoration of function (14).

Modification and Splicing of mRNAs

During cleavage and processing of primary RNA transcripts, RNA molecules are capped at their 5′ origin by the addition of methylated G nucleotides, and a segment of poly-A residues is attached near their 3′ end (Fig. 14-1). These modifications are necessary to allow export of the transcript from the nucleus, to stabilize mRNA molecules, and to allow interaction with ribosomes. The presence of a poly-A tail is a major characteristic of mRNAs and constitutes a key tool to isolate mRNAs from other RNA molecules, mostly ribosomal, using complementary oligo-dT primers.

Before RNA is exported into the cytosol, splicing of intron segments is performed. As introns facilitate recombination events, they play a major role in the evolution of species and are often targeted in the process of generating knockout animals. Their physical structure may be more important than their actual nucleotide sequence, as these diverge more rapidly between species than exons (15, 16). Introns can range in size from 80 to several thousand nucleotides. They contain specific sequences at their extremities referred as the 5′ splice site (always ending with 5′-GU-3′) and the 3′ splice site (always ending with 5′-AG- 3′) that come together in the process of splicing. During splicing, a large catalytic heteromeric complex, termed spliceosome, is formed by the assembly of different ribonucleoproteins (1, 2). After RNA binding, spliceosomes bridge together two exons and excise their intron segment. Generally, each 5′ splice site pairs with the closest 3′ splice site in the spliceosome, producing only one form of mRNA. In some cases, however, splicing of RNA enables a single gene to produce different mRNA transcripts by jumping from a given 5′ splice site to a more distant 3′ splice site (Fig. 14-1). Each of these mRNAs yields different isoforms of the same protein that can be alternatively produced in different types of cells. Genome-wide analysis of

alternative splicing indicates that 40 to 60% of human genes have alternative splice isoforms (16, 17). Splicing of mRNA is an important regulatory step in the production of cell proteins that requires high degree of accuracy. In general, DNA mutations that involve splice sites (splice site mutations) do not prevent splicing but instead cause the normal partner to seek alternative splice sites, producing the synthesis of various abnormal proteins lacking one or more exons, as, for example, in Frasier syndrome (18). Constitution of gene banks containing thousand of gene transcripts has greatly facilitated the recognition of splicing variants of a given gene, which can be easily accessed online: (www.ncbi.nlm.nih.gov/IEB/Research/Acembly/).

Regulation of Proteins Synthesis

Mammalian ribosomes are the site of mRNA translation and composed of two asymmetric subunits - the 40S that binds mRNAs and the 60s that interacts with transfer RNAs. Eukaryotic cells can decrease their rate of protein synthesis in various conditions such as infections or heat shock. One important mechanism that mediates these types of nonselective responses involves phosphorylation of a repressor protein termed elF-2 that interacts with target regions containing the AUG start codon preventing ribosomal binding. A second gene-specific mechanism of translational regulation is termed attenuation and involves the formation of mRNA hairpins that block the translation. Similar mechanisms also regulate mRNA stability and degradation by RNases. Given the fact that each molecule of mRNA can serve as a template for multiple copies of proteins, the rate of mRNA degradation is a major determinant of protein abundance and is often the site of complex regulatory processes. In general, these processes operate primarily on unstable mRNAs (such as cytokines) that are stabilized under specific conditions by their interaction with *trans*-acting elements (19, 20). The stability of transferrin mRNA, for example, increases during iron deprivation, triggering the synthesis of more transferrin molecules (21).

Protein Sorting and Degradation

Newly synthesized polypeptides are processed by a complex network of cellular enzymes and other binding proteins that are arranged in a highly organized fashion in various organelles in the cell. Information resident in the primary amino acid sequence as well as the folded structure of the proteins allow each to be recognized and targeted to its ultimate destination (1, 2). Proteins which are synthesized in free cytosolic ribosomes are normally directed to the nucleus or the mitochondria, whereas membranous ribosomes are the site of synthesis for proteins that enter the reticuloendothelial system (co-translational transport) to be redirected to their final destination after being processed in the Golgi apparatus. An excellent example of this process is highly polarized epithelial cells in renal tubules. In these cells, transport proteins are located specifically on the apical or basolateral plasma membranes. This arrangement enables epithelial cells to perform net transport of solutes and water to either secrete or reabsorb fluid. Some proteins are able to self-assemble by spontaneous interaction among reactive amino acid groups, whereas other proteins, such as the V2 receptor, require assistance of molecules such as the Hsp70 system and chaperonins (22). These molecules control accessibility of reactive groups and maintain the peptide in a relatively flexible state until it reaches its final conformation. The final fate of most cell proteins is degradation into proteosomes. A process called ubiquitination that involves covalent linkage of small peptides called ubiquitins to target proteins precedes this step. Ubiquitination is also involved in important signal-transduction pathways, such as the nuclear factor- κB pathway - in which inhibitory subunits are degraded after stimulatory signals and activate various signaling cascades (23).

Recombinant DNA Technologies

Introduction

Until recently, molecular biology has addressed two major objectives; namely, to identify genes and analyze their function. With the completion of the human genome project, focus is shifting from the first to the second goal and to the more complex task of delineating complex patterns of gene expression. In the following sections, the more common recombinant DNA and protein analysis technologies are described, to illustrate common experimental procedures that are routinely used in molecular diagnostics as well as basic research.

Basic Recombinant DNA Technology

Hybridization and Detection of Nucleic Acids

Pairing of nucleotide bases in DNA and RNA allows a wide variety of specific recognition processes both *in vivo*

and *in vitro*. These not only form the basis of many critical cellular functions, but also provide the molecular biologist with tools to detect and study single genes.

The extraordinary specificity of nucleic acid-base recognition has been exploited in the process of hybridization of complementary DNA or RNA *in vitro*. Under appropriate conditions, a unique nucleotide sequence present within a complex mixture of nucleic acids can be identified with a resolution of greater than one part per million. Using standard techniques, DNA or RNA is isolated from cells or tissues, stripped of proteins, and denatured into single strands. When incubated under conditions favoring renaturation, complementary sequences reassociate. Experimental conditions (commonly temperature and salt concentration) can be altered to allow for only perfect or nearly perfect sequence matches. This is called stringency. Lowering stringency conditions is sometimes desirable to identify close relatives of particular nucleotide sequences. Thus, one can search for a gene or mRNA in kidney that is a close relative of a transcript expressed in other tissues as well as species. Nucleic acids are commonly fractionated by agarose gel electrophoresis. Under these conditions, the agarose acts as a molecular sieve, retarding larger strands while allowing smaller strands to migrate in the electric field placed across the gel. Fractionated DNA or RNA is then eluted from the gel or can be transferred to a filter and exposed to a labeled DNA probe. When the filter contains DNA, this process is called a Southern blot, whereas it is called Northern blot if it contains RNA (1, 2).

This same procedure has also been adapted to tissue sections to localize expression of mRNA transcripts by specific cell types within a complex organ such as the kidney. This technique is called *in situ* hybridization (1).

Restriction Endonucleases

The discovery by Arber in 1962 of bacterial nucleases that cut DNA molecules at specific locations constitutes one of the cornerstone in the development of recombinant DNA techniques (24).

Most restriction enzymes recognize palindromic DNA sequences, meaning that the 5′ to 3′ sequence in the upper strand is identical to the 5′ to 3′ sequence in the lower strand.

To date, nearly 1,000 different restriction enzymes have been purified. Each enzyme can produce defined DNA restriction fragments possessing specific nucleotide sequences at each end from any given DNA sample. With these enzymes, strings of DNA can be isolated, ligated into plasmid or phage genomes, and amplified (Fig. 14-3).

Historically, restriction nucleases have allowed construction of the first detailed maps of various genomes (restriction maps). They are commonly used for allelic discrimination by restriction fragment length polymorphism (RFLP). In this technique, DNA mutations which modify the recognition sequence for specific restriction enzymes can be identified by the length of the restriction reaction product.

DNA Amplification Using Procaryotic Systems

The possibility of replicating specific strings of DNA to obtain quantities sufficient for analysis and further manipulation is central to all recombinant DNA technologies. In the early 1970s, work by Boyer and by Cohen provided the first fundamental tools for DNA cloning with the discovery of DNA ligases and the characterization of bacterial plasmids.

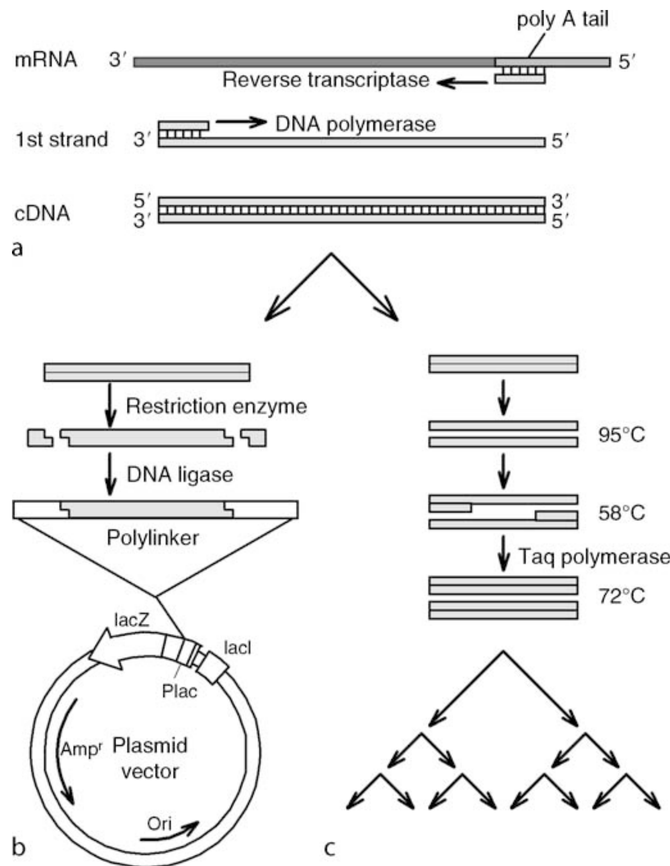
Plasmids are circular molecules of DNA that replicate in the cytoplasm of bacterial cells (Fig. 14-3). Specific regions of plasmids not vital for vegetative growth under laboratory conditions can be engineered using restriction nucleases for insertion of exogenous DNA fragments. When inserted into *E. Coli* cells, plasmid genes encoding for antibiotic resistance are expressed, allowing selection of bacteria that have been transformed. During this process, the inserted DNA is replicated along with the rest of the bacterial genome. This technology enables amplification and characterization of virtually any DNA string of appropriate size.

Because bacteria has a mean generation time of approximately 20 min during exponential growth and contains up to several hundreds copies of plasmids per bacterium, virtually limitless amounts of DNA can be grown and harvested routinely. Most available plasmids are engineered to incorporate a polylinker corresponding to a portion of DNA that contains multiple restriction sites that facilitate cloning of DNA fragments generated by various restriction enzymes.

Because of their simplicity, DNA is generally cloned into plasmids. The relatively low efficiency of bacteria transformation with plasmids limits their use however, when generating DNA libraries. In this case, bacterial viruses such as the *bacteriophage* λ, are more advantageously used.

The λ phages are composed of a head that contains the viral genome and a tail that infects bacterial cells with high efficiency. Phages can be packaged *in vitro* after insertion of exogenous DNA and are replicated in

Figure 14-3
Basic enzymes and techniques in recombinant DNA technology. The figure summarizes basic procedures used in routine laboratory experiments. a. mRNA molecules are reverse-trascribed into cDNA using enzyme reverse transcriptase in order to synthesize the 1st complementary DNA strand, which is then used to generate double-stranded complementary DNA (cDNA) using a DNA polymerase. Both enzymes require prior priming with complementary oligonucleotides. b. cDNA and other DNA molecules can be amplified in bacteria after insertion into plasmids using restriction enzymes and DNA ligase. Recombinant plasmids generally contain a polylinker that offers different restriction sites to facilitate insertion of exogenous DNA. In addition, they are engineered to contain antibiotic resistance genes (Ampr) for selection of transformed bacteria and often contain other useful genes like the lacZ in this example, which allows color detection of colonies transformed with "empty" plasmids. c. Alternatively DNA can be amplified by PCR with the Taq polymerase. Repeated cycles of DNA denaturation (95°C), primer annealing (58°C in this example), and DNA synthesis (72°C) allow for exponential replication of DNA strands encompassed by the two primers. When reverse-transcription and PCR are combined, the procedure is referred to as RT-PCR.



bacteria. Both λ phages and plasmids can amplify DNA fragments up to 20 kb. They are ideal for cDNA (see below) and other relatively small DNA molecules, but are insufficient for large strings of genomic DNA. In these cases, other vectors can be used. These include cosmid vectors, containing elements of both plasmids and λ phages that can accommodate up to 45-kb fragments or bacteriophage P1 housing up to 100 kb of exogenous DNA. If larger fragments need to be replicated, plasmid vectors based on the *E.Coli* F vector can incorporate up to 300 or 1,000 kb of DNA. The bacterial artificial chromosome (BAC) is a derivative of the F plasmid and is present in very few copies per *E.Coli* cell. Currently, BACs are the preferred vectors for genomic DNA libraries (1, 2). Yeast artificial chromosomes (YAC) can also be used for the same purpose.

DNA Amplification with Polymerase Chain Reaction

The second breakthrough in DNA amplification was achieved in 1985 by Mullis and coworkers who developed PCR (25).

PCR permits selective amplification of minute quantities of DNA, facilitating every aspect of molecular biology research and diagnostics including site-directed mutagenesis, labeling, and sequencing of DNA. Fixed tissues on slides or small tissue fragments, such as renal biopsy specimens, can also provide sufficient material for PCR amplification.

PCR reaction has nowadays replaced Southern and Northern blotting in most recombinant DNA applications.

PCR relies on the binding of two chemically engineered priming oligonucleotides (primers) that flank a region of DNA, to amplify the region located in between (*Fig. 14-3*). Primers are complementary to the opposite DNA strands. Addition of DNA polymerase results in the synthesis of new DNA. Repeated cycles of denaturation, annealing, and DNA synthesis are performed in a chain reaction such that the newly synthesized strands become templates for further DNA synthesis. This process exponentially increases the number of DNA copies containing the sequence of interest. This allows isolation of a given DNA string from the genomic DNA or from a pool of cDNA molecules (see below).

PCR reactions are performed using thermostable DNA polymerase species derived from the thermophilic bacterium Thermus aquaticus (Taq polymerase), which retain activity after being heated to 95°C. A number of modified enzymes that guarantee reliable DNA duplication or amplification of long DNA strings are commercially available.

Since the complete sequencing of the human genome, *in silico* PCR can now be performed to identify PCR products.

Amplification of mRNA with Reverse Transcription

In several experimental circumstances direct amplification of mRNAs is required.

mRNA transcripts reflect the actual genes that are activated in a cell system and contain nucleotide sequences which can be directly translated into proteins, obviating the tedious task of sorting exon segments from introns when working with genomic DNA.

This process is achieved with a reverse transcriptase derived from retroviruses, which is one major exception to the central dogma, as it harbors their genetic information in RNA molecules that are copied into DNA on infection of host cells.

The DNA obtained by reverse-transcription is called cDNA; it reflects the nucleotide sequences of mRNAs (*Fig. 14-3*). Similar to DNA polymerases, reverse transcriptase requires complementary oligonucleotide priming to begin transcription. Oligo-dT hybridizing to poly-A tails or random primers can be used to reverse-transcribe mRNA molecules (3, 9) in a non-selective manner.

Once converted into cDNA, nucleic acids can be ligated into vectors or directly amplified by the PCR reaction, a process that is referred as reverse transcription polymerase chain reaction (RT-PCR) (*Fig. 14-3*). In other circumstances, gene-specific primers are designed to amplify selected mRNA molecules.

As investigators are progressively turning their attention from genomic analysis to the analysis of gene expression, this process has become extremely important. It permits the generation of collections of "Expressed Sequence Tags" (ESTs), which provide extremely rapid tools to identify genes, evaluate their expression, and construct genome maps. ESTs are small DNA sequences that are generated by RT-PCR reactions and which when pooled together, represent a collection of DNA sequences that are expressed in a given cell or tissue. These "tags" can be used to identify specific portions of the genome that encode for a given gene, "fish-out" similar genes, or identify splicing variants of a given mRNA, using online collections of ESTs.

Sequencing Nucleic Acids

The highly specific binding of small oligonucleotides to DNA also lies at the heart of the dideoxy chain termination sequencing of DNA.

DNA sequences are obtained from a uniform population of DNA, obtained by PCR.

After DNA is denatured, complementary primers are added. In the traditional manual sequencing, synthesis of complementary radioactive DNA strands was initiated by DNA polymerase after addition of ^{35}S-labelled radioactive deoxy-nucleotides in four different reactions containing one of the four dideoxy-nucleotide analogs of G, A, T, or C. In each reaction, chain termination occurs if a dideoxy analog is inserted in the newly formed DNA strand, preventing further extension.

Modern automated sequencing machines apply these basic principles using fluorescence labeled deoxynucleotides. The reaction products are resolved on sequencing gels or electrophoresis capillary. As DNA advance along the electric gradient, a laser beam is used to excite their fluorescence, which is read and analyzed by a computer. Reliable DNA sequences can generally be obtained for more than 500 nucleotides per run, and multiple lanes can be read simultaneously.

Analysis of Gene Expression

A critical issue in normal physiology and renal pathophysiology is the determination of the expression of given genes under different cellular and environmental conditions.

Classically, gene expression analysis is performed by protein detection with specific antibodies in Western blotting or, when antibodies are not available, by measuring the amount of mRNA transcripts. Densitometric methods have been developed to compare the amount of expressed protein or mRNA, with respect to a control preparation. These semi-quantitative techniques however, have severe limitations. They require relatively large amounts of starting material, can only study a limited numbers of genes simultaneously, and require development of specific probes such as antisera. To overcome some of these limitations, other techniques have been developed and are briefly reviewed.

Real-Time RT-PCR

Unquestionably, RT-PCR is more sensitive than the traditional Northern blot analysis to measure levels of mRNA expression, and can be performed from limited amounts of mRNA.

The major difficulty in quantifying mRNA by RT-PCR however, is related to the exponential nature of the method, which tends to amplify differences when comparing levels of expression in different biological conditions or biological systems.

For this reason, competitive RT-PCR was initially developed and was based on the use of internal standards sharing the same priming sequences as the transcript of interest, which were added to the mixture to act as competitors during the reaction (26). This method was however time consuming and has now been replaced by real-time PCR and micro-array analysis when quantitative expression of multiple genes needs to be evaluated.

Real-time RT-PCR allows detection of PCR products as they are being formed (*Fig. 14-4*), using quenched fluorescent dyes linked to the 5′ end of one primer that are released by the 5′ nuclease activity of the Taq DNA polymerase. The emitted light is measured in real time during PCR reaction and is proportional to the amount of PCR product. The number of cycles required to cross a given fluorescence threshold is inversely proportional to the amount of mRNA present in the original reaction mixture. Multiplex real-time RT-PCR represents an extension of this technique and is based on differential fluorescent labeling of primers that amplify for different genes. This permits comparison within the same PCR reaction of the relative amount of up to 3–4 different transcripts (27). Generally, a house-keeping gene that is presumed to be stably expressed, serves as an internal control, allowing correction of the results for the amount of RNA that was loaded in the initial sample reaction.

Multiplex real-time RT-PCR allows determination of gene expression even from extremely small samples such as renal biopsies - and has been developed for example, to detect gene expression in human renal specimens (28) or quantify viral genome copies in biological samples.

Differential mRNA Display Using DNA Microarrays

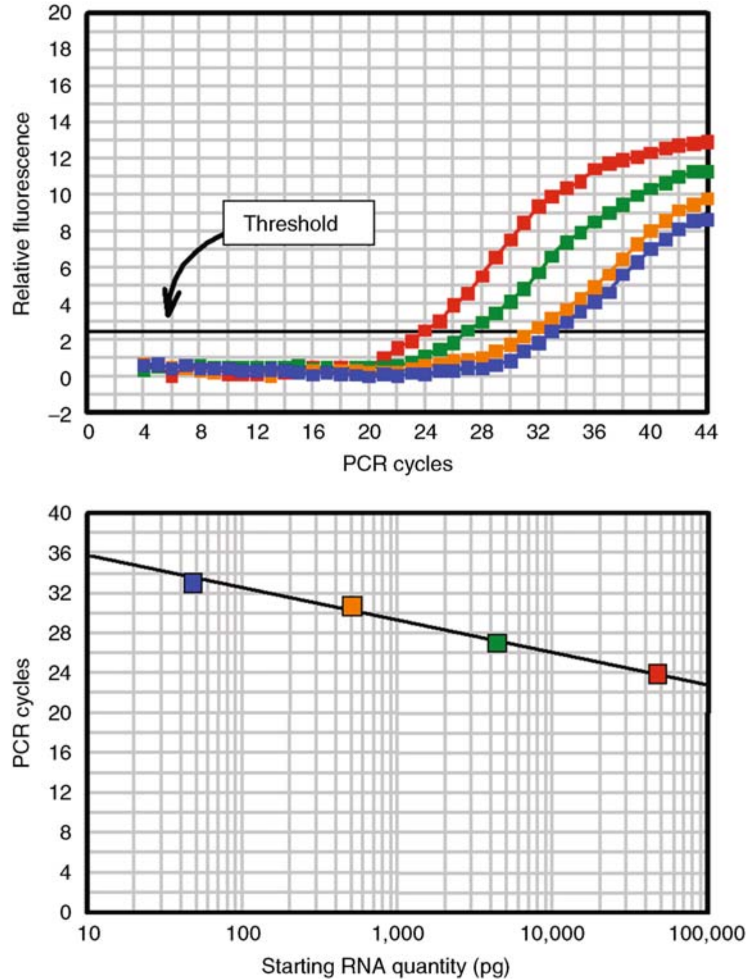
DNA microarray technology has become a valuable technique for comparative gene expression analysis. Transcriptomic DNA chips are composed of thousands of known expressed sequence tags or synthetic oligonucleotides, which are deposited in gridded arrays by a robotic spotting device on a solid support such as a glass microscope slide or a membrane matrix (29).

Oligonucleotide sequences are selected from databases such as GenBank, dbEST, or UniGene (30). Several thousand genes can be spotted on a single microscope slide for large screening. Other chips contain clusters of genes that are functionally related, as in oxidation chips or cell-cycle chips, for example. Two populations of mRNA are labeled with different fluorophores, hybridized to the same chip, and analyzed with a laser scanning device (*Fig. 14-5a*).

The intensity of the fluorescence emitted by each dye is proportional to the amount of RNA that has hybridized at a given location, reflecting the level of a gene expression represented by that spot.

Despite increasing developments, this technique has limitations, particularly in terms of chip reproducibility and variability in the efficiency of labeling and hybridization, which generally need to verify the obtained results by standard RT-PCR or real-time RT-PCR.

Figure 14-4
Real-time PCR. Figure illustrates an actin calibration curve. Total RNA was extracted from HK2 cells and loaded in increasing concentrations in the sample reaction. Fluorescence was measured in real-time as primers were incorporated in newly synthesized PCR fragments with an ABI Prism 7,700. The number of cycles required to cross a given fluorescence threshold shown in the upper panel is proportional to the initial amount of loaded mRNA as shown in the lower panel.



The enormous quantity of information generated by expression data from thousands of genes requires sophisticated computer analysis to generate meaningful results (31). Computer-based algorithms have been developed to recognize patterns of gene expression within complex genetic networks such as the human genome. The so-called cluster analysis is a powerful statistical tool, permitting grouping of genes in hierarchical clusters that follow similar patterns of expression (*Fig. 14-5b*). This information can then be used as a molecular fingerprint for diagnosis or monitoring response to therapy. It may allow detection of subtle changes of gene expression and may identify functions and interactions of uncharacterized proteins, by grouping them into clusters of genes whose function is known.

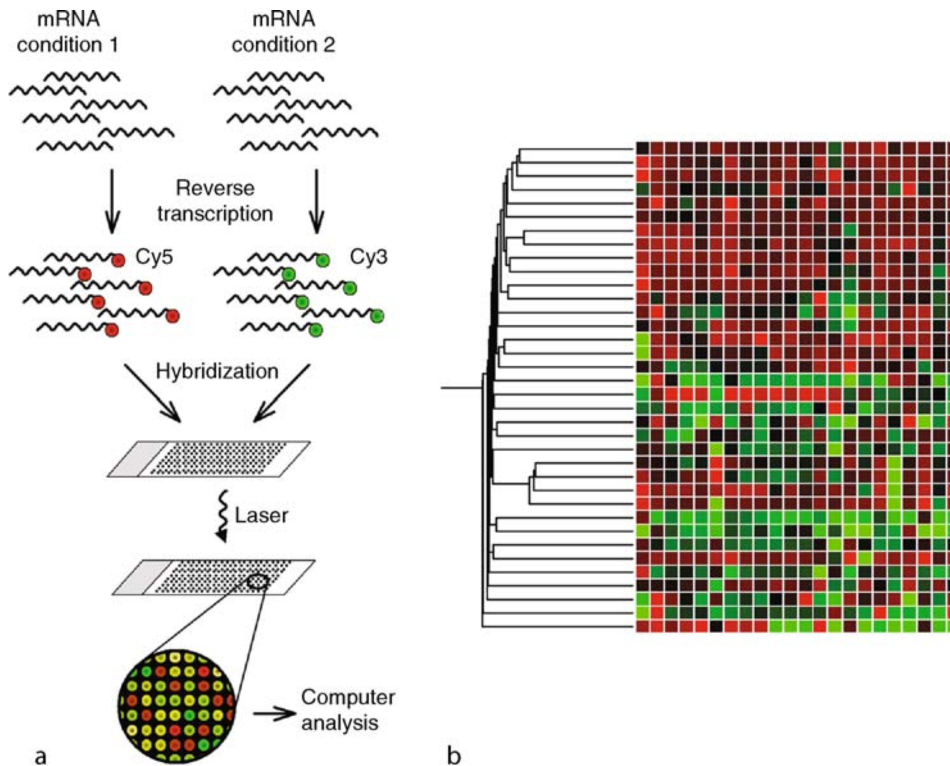
Microarray Technologies for Genomic Analysis

Similar approaches to the transcriptomic chips have also been developed for the analysis of genomic sequences.

In “comparative genomic hybridization” (CGH) for example, genomic gain or loss of particular genes can be detected, permitting identification of small deletions or duplications throughout the genome.

Figure 14-5

Differential display on DNA microarrays. (a) mRNA obtained from two different samples are labeled with two fluorescent cyanine dyes with one round of reverse-transcription. The fluorescent targets are then pooled and hybridized under stringent conditions to the clones on the microarray chip. The emission light is measured with a scanning confocal laser microscope at two different wavelengths that are specific for each dye. Monochrome images are then pseudo-colored, combined, and analyzed with a suitable computer software. (b) Cluster analysis of results obtained in a microarray experiment. Rows represent individual genes whereas columns indicate different experiments. Genes that are the most up-regulated are colored in red, while down-regulated genes are indicated in green. By analyzing patterns of gene expression, genes that behave similarly are grouped together in hierarchical order, to identify patterns of gene expression.



Microarray chips have also been developed to detect gene mutations. In this case, the target is limited to few genes. Sequential small fragments of the genes of interest are arrayed onto the chip, which detects even single nucleotide changes, when hybridized with amplified DNA obtained from patients carrying the mutation.

Likewise, microarrays can also be used to screen for single nucleotide polymorphisms (SNPs). Chips containing over 1.6 SNP markers have now been developed allowing whole-genome genotyping to search for associations of particular SNPs with given clinical conditions. Using this approach for example, genes that are involved in the pathogenesis of Systemic Lupus Erythematosus have been identified (32), and whole genome association studies can be done with large populations.

Gene Cloning and Analysis of Cloned Sequences

Until recently, one of the primary goals in molecular research was to clone genes responsible for diseases. This task has been largely accomplished for monogenic diseases, although several genes, mostly encoding for proteins involved in rare diseases, still need to be identified. Other diseases have a polygenic base of inheritance and a vast majority of these that regulate complex patterns of inheritances still need to be identified.

Cloning of genes responsible for genetic disorders can be achieved without prior knowledge of the molecular nature of the disease by a strategy termed positional cloning, which is based on the genomic localization of

the locus of interest using genetic markers present at a known chromosomal location.

Alternatively, gene cloning has been performed by screening tissue-specific collections of recombinant vectors containing sequences of cDNA, termed cDNA libraries. This basic approach can be refined by the use of subtractive libraries for example, that are obtained by subtracting unwanted mRNA species from the library before screening.

Currently, the process of cloning genes has been completely revolutionized by the creation of electronic databases. More than 100,000 partial ESTs sequences that cover most human genes have been collected (33), and thousand of putative genes that were generated by digitally sorting out and mending together potential exon regions of the human genome have been obtained.

These sequences are collected in databases that can be accessed online, which constitute virtual DNA libraries to be screened electronically using partial DNA or protein sequences. Tags sequences can be obtained from differentially expressed proteins or DNA molecules (identified by microarray analysis for example) or from sequences of related genes. Once a putative gene has been identified, its sequence can be directly amplified from mRNA or from genomic DNA, by RT-PCR or PCR, respectively. As already mentioned, ESTs databases also represent invaluable tools for "serial analysis of gene expression" (SAGE), which is considerably increasing our knowledge of the human transcriptome (33).

Other techniques that are used to identify genes include library screening with antibodies, functional assays, or by protein-protein interaction. Using expression plasmids, peptides encoded by exogenous cDNAs can be directly translated in the bacterial lawn and identified with specific antisera (34). With a similar approach, phage epitope libraries expressing randomly generated oligopeptides can be screened to find domains recognized by antibodies or by other proteins (35). The recognized epitope sequence generally corresponds to a partial amino acid sequence contained in the natural antigen or binding protein. From this sequence, the protein can be identified or cloned. This technique is particularly powerful for identifying autoantibodies or for cloning proteins by their reciprocal interactions such as in receptor-ligand association.

The two-hybrid system is an alternative strategy which permits investigators to fish for clones which code for peptide that interact with other proteins offered as bait in the screening process (1, 2).

The strategy of expression-cloning relies on screening cDNAs that produce functionally active proteins when expressed in a suitable system, such as Xenopus oocytes. Assays using oocytes are usually used to identify and study by electrode impalements, patch clamping, and flux techniques proteins involved in membrane transport. Large quantities of RNA (cRNA) can be synthesized *in vitro* and injected into oocytes (36).

Online programs provide important clues regarding the nature of newly cloned cDNAs and proteins, such as structural aspects, including membrane-spanning domains or antigenicity, presence of specialized amino acid sequences coding for functional domains such as phosphorylation, glycosylation, or targeting domains to specific cell compartments (www.ncbi.nlm.nih.gov/IEB/Research/Assembly/). Sequence alignment in databases helps define relationships with other genes and identify functional motifs, such as DNA-binding and protein-binding domains, that can give important clues to the biologic function of the newly cloned sequence (37). In addition, important information is also contained in the promotor region that can be identified and screened for consensus sequences that help defy the physiological role of the protein in the cell.

Gene Expression and Silencing in Cell Cultures (see also Chapter on In Vitro Metods in Renal Research by Dr. Wilson in this text, Chapter 15)

One important aspect of recombinant DNA technology is the demonstration of the biological role of a selected gene. The easiest and often first approach, is to express or suppress a given gene in cells that are cultured *in vitro*.

Cells can be obtained from tissues after disruption of the extracellular matrix with enzymes, such as trypsin or collagenase. Cells of the same type need thereafter to be purified using different techniques that are based on cell size, resistance to specific conditions, or expression of specific markers. In the latter case, cells are often isolated by fluorescence activated cell sorting (FACS) or with beads that bind to specific antigens. In highly organized tissues, such as the kidney, laser captured microdissection can help isolate fragment of tissues containing only few cell types, which greatly facilitates the following purification steps. These "primary cell cultures" are extremely powerful biological models, because they often retain much of their original phenotype. Unfortunately, they often grow slowly, their preparation is expensive and time consuming, and they tend to stop growing after few passages due to a process termed "replicative cell senecesce". This process is in part caused by the lack of telomerase, which prevents shortening of telomeres at each cell division. Introducing the catabolic sub-unit of the telomerase gene allows in some cases to obtain "immortalized cell lines". In most cases however,

mammalian cells stop dividing as they activate cell-cycle "check-point mechanisms". In order to inactivate these mechanisms, viral pro-oncogene genes are usually inserted to generate "transformed cell lines". These cells proliferate indefinitely, but unfortunately tend to loose their phenotype. To partially circumvent this phenomenon, strategies aimed at turning-off the pro-oncogene can be used. The thermo-sensitive Simian Vacuolating virus 40 T antigen (SV40 Tag) for example, promotes undifferentiated cell proliferation at 33°C, but is turned-off at higher temperatures. Conditionally immortalized cell lines can therefore be grown to confluence at 33°C, but will stop proliferating at 37°C and will, in most cases, recover part of their original phenotype (38).

To study the effects of a given gene, cell cultures and cell lines can be directly obtained from specimens of patients lacking a given gene. Alternatively, genes can be over-expressed or suppressed in cell cultures. Genes are usually inserted into expression vectors under the control of a potent viral promoter, like the CMV promotor. Cells are thereafter "transfected" with these vectors. Transient transfection allows for gene expression for a few days. Stably transfected cell lines can also be obtained by transfecting cells with a linearized vector that will insert itself randomly into the cell genome in few cells. As most of these vectors contain an antibiotic resistance gene, stably transfected cells can be selected and purified.

Vectors harboring mutated genes can also be engineered by site-directed mutagenesis techniques, which allow selective mutatation, deletion or insertion of peptide residues of a given protein. Once expressed, these mutated peptides produce information on the function of various domains of a single protein (39).

Specific cDNAs sequences can also be fused to sequences encoding for fluorescent proteins (GFPs). Once, these vectors are transfected into cells, it will prompt the synthesis of a fusion protein which is composed of the protein of interest and a GFP tag, which helps to follow the protein expression in the cell by fluorescence microscopy.

Similarly, gene regulatory regions such as promoters can be transfected. Their effects on a neighboring reporter gene can then be determined using a gene product that is easily assayed, such as chloramphenicol acetyltransferase or a luciferase.

In some cases, non mammalian cells are more advantageously used. *Xenopus* oocytes are popular systems to express membrane transport proteins. The cystic fibrosis gene was among the first gene to be isolated without knowledge of its actual function. Oocytes were used to demonstrate its function as an epithelial cell chloride channel (40). Other fundamental membrane transport proteins have similarly been cloned using Xenopus oocytes, including Na-H exchangers (41), bumetanide-sensitive Na-K-2Cl and thiazide-sensitive NaCl cotransporters (42), the renal outer medullary adenosine triphosphate–regulated potassium channel (43), multiple aquaphorin water channels (44), and the amiloride-inhibitable epithelial Na + channel, or ENaC (45).

These studies have provided molecular links between epithelial cell transport data and the expression of specific genes within individual kidney epithelial cells.

Detailed knowledge of these transporter proteins has also enabled the identification of specific gene abnormalities in humans that cause inherited disorders of renal tubular function including nephrogenic diabetes insipidus (46) and Bartter's (47, 48), Gitelman's (49), and Liddle's (50) syndromes.

Although gene suppression in animal model is probably the most powerful approach to study the function of a given gene, another approach is to inactivate its mRNA. This is achieved by a technique called RNA interference (RNAi, siRNA), in which double stranded RNA molecules matching the sequence of interest are introduced into cells, where they hybridize with their complementary mRNA, causing its degradation. Fragments of degraded RNA form other double-stranded RNA, which continues to eliminate more targeted mRNA. In addition, some RNA molecules enter the nucleus where they inhibit directly gene transcription by interaction with the targeted genomic sequence (51).

The range of applications of recombinant DNA technologies to protein expression expands well beyond the above mentioned techniques. Bacteria for example, can also be "transformed" with recombinant plasmids containing bacterial promoters that activate transcription of genes fused to specific detection sequences. These fusion proteins can then be purified in large quantities and used for functional studies or as immunogens to raise antisera. As mammalian proteins expressed in bacteria are not post-translationally modified, viral expression systems have been developed, allowing the production of recombinant proteins by viral infection of cultured insect cells (52). These proteins are then properly processed, glycosylated, and phosphorylated. A similar approach can be used in yeast. In fact, yeast cells have in fact become particularly interesting for the study of the cellular effect of given proteins. The 6,000 yeast genes have been fully sequenced, are particularly easy to mutate, and large collections of well characterized mutant strains are available. This is enabling researchers to perform functional genomics and proteomics studies in a simple organism and has

enabled dissection of genomic control mechanisms as well as identification of several proteins that regulate a variety of cell functions, including endocytosis and membrane fusion (53).

Expression and Suppression of Specific Genes in Animal Models

In vitro cell expression systems are very powerful tools, but have limitations when studying the role of genes in multicellular organisms in which a gene's expression or lack thereof has complex effects on an animal's development and physiology. In these cases, genetically engineered animals which are modified in genes that are homologous to their human counterparts, are used advantageously. Whenever the animal phenotype is similar to the human disease, it always represents an extremely powerful tool to understand the disease and test new treatments.

Likewise, scientists have studied genes which are responsible for different animal phenotypes or have been genetically manipulated to modify activity of given genes, for which the human homologue has not been associated with specific diseases. Most of these experiments have been performed in mice, but also in other organisms such as *S. Cervisiae, C. Elegans, Arabidopsis,* and *Drosophila*. This has led to the constitution of a collection of animal mutations that represent invaluable repertoire of candidate genes for human diseases which can be tested based on clinical phenotypes. This approach is referred as "reverse genetics", because instead of identifying a given gene using biological material obtained from patients, scientists begin their search from experimental gene mutation data, to identify by phenotypic homology human diseases.

While engineering animal models of a given disease, different strategies are used depending on the effects of the human mutations ("loss of function" or "gain of function") and the mode of inheritance (recessive, addictive or dominant).

In most cases the removal of a given gene, termed "knock out", is the first approach to reveal the function of its encoded protein. Other strategies are aimed at changing the levels of expressions of a given gene or changing its expression in specific tissues or in time. This latter approach, which is generally based on the use of inducible promoters, is particularly interesting when analyzing genes that are implicated during development or when the mutation is lethal in animals. In this case, the gene of interest can be "turned-off" only after birth, when the animal is fully developed. In some cases, researchers have adopted a dominant-negative strategy, when over-expression of a mutant protein can inhibit by competition, the activity of its wild-type homologue, which continues to be normally synthesized. In general, gene replacement or addition is more complicated and time-consuming than gene knock-out. Regardless of the strategy that is used, all these genetically modified animals are termed "transgenic" and their artificially modified genes are referred as "transgenes".

Introduction and disruption of specific genes into amphibians and insects such as *Xenopus* and *Drosophila* are particularly useful in the analysis of developmental genes and have been used extensively to characterize various developmentally specific transcripts governing tissue-specific differentiation, including in the kidney (1, 2). This research is greatly facilitated by the ability to manipulate easily, cells of the earliest embryo stages and have produced a fundamental understanding of pattern formation in these animals.

For most human diseases, mice have become the animals of choice, because their genes can now be easily manipulated, their genome has been fully sequenced, a full range of techniques have been developed to analyze their phenotype and they can be rapidly bred to produce heterozygous and homozygous mutants or compound transgenic animals, in which more than one gene has been modified (54–56).

In some cases, production of transgenic animals may be performed by injection of the transgene directly into the pronucleus of a one-cell stage embryo so that it can integrate into the genome. In mice, a vector carrying the transgene is introduced in embryonic stem cells (ES) which are allowed to proliferate *in vitro.* The rare cells where homologous recombination with the original gene has occurred are then selected, and injected with a micropipette into the embryo at its early stage. This leads to the formation of a chimeric animal that will carry the mutation in a significant percentage of its germ lines. Mice are then bred to produce heterozygous male and female off-springs, which, once mated together, will produce homozygous animals. Both heterozygous and homozygous animals can then be studied.

Conditional mutants allow for the disruption of genes in specific tissues at given times.

To express a specific protein in podocytes, for example, the nephrin promotor can be used, in order to activate the transcription of a given gene only in cells that can activate this promotor (56). Using this strategy for example, researchers have over-expressed the macrophage migration inhibitory factor (MIF) gene in podocytes, demonstrating the development of mesangial sclerosis in the presence of high levels of MIF (57).

To knock-out genes at a given time, site-specific recombinant systems are used, like the Cre/Lox system. For this purpose, a fully functional gene or a portion of this gene is flanked by small sequences of DNA corresponding to the "lox" sites, which are recognized by the Cre recombinase protein. Transgenic animals are then mated with mice expressing the Cre recombinase under the control of an inducible promotor that excise the gene of interest, when activated (56).

Protein and Peptide Analysis

Principles and Techniques

The principle of protein analysis has considerably evolved over the past years in parallel with key technological advancements, especially in the field of mass spectrometry.

The basic principle in proteomic analyses is that the digestion products of a given peptide create a fingerprint of the original protein, which permits its identification.

In most cases, identification of a protein requires a preparative step, to obtain a purified sample. Two-dimensional (2D) electrophoresis is frequently used for this purpose, while LC-Mass and MALDI-TOF are commonly used for the analysis of digestion products.

Denaturing and Non Denaturing 2D-Electrophoresis

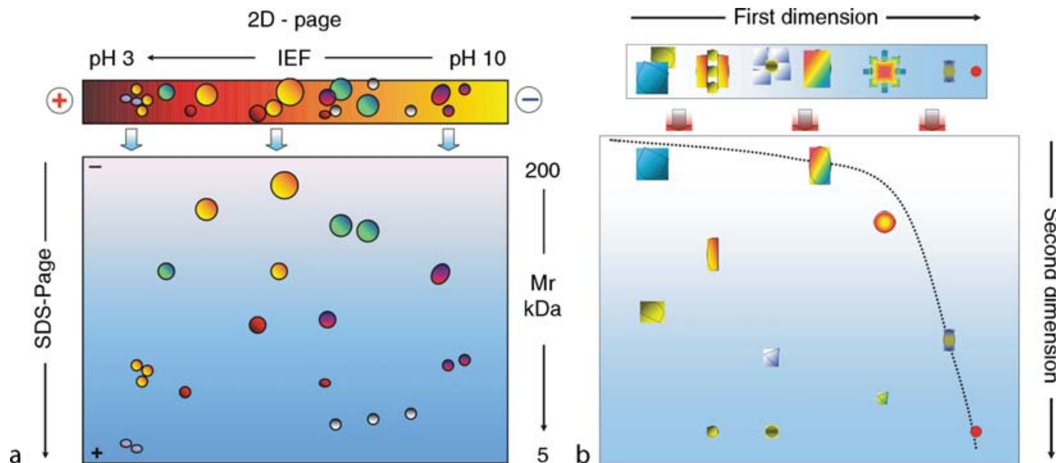
Classic 2D protein electrophoresis is the technique of choice for the analysis of plasma proteins or other complex biological samples. High-resolution is usually achieved by combining separation by charge and separation by size in denaturing conditions (IEF/PAGE) (*Fig. 14-6*) (58, 59). This can then be followed by micro-scale mass spectrometry that allows identification of individual spots. The development of soft gels has considerably improved the resolution of high molecular weight proteins, but requires denaturing conditions, which prevents the analysis of protein-protein interactions (*Fig. 14-6b*) (60). For this reason, new techniques that separate protein mixtures in low denaturing conditions have recently been developed. These include Blue-PAGE, which is performed on membranes, or the Nat/SDS PAGE, which is performed on a polyacrylamide substrate (61–63). This latter system helps to resolve protein aggregates and disclose protein interactions (*Fig. 14-6b*).

Protein Staining

Traditionally protein staining after electrophoresis has been performed with Coomassie R-250 and silver ions. New dyes allow differential protein expression analysis on

Figure 14-6

Schematic representation of a classical 2D-polyacrylamide gel electrophoreis (2D-PAGE) (a) and of a 2D electrophoresis in non- denaturing condition (Nat/SDS-PAGE) (b). In the former approach, proteins are first separated according to their charge in the presence of urea (IEF) and are then run in a polyacrylamide gradient that separates them on the basis of size. In Nat/SDS-PAGE protein complexes migrate unresolved in the first dimension and are then separated in denaturing conditions, in the second dimension.



2D-gels (DIGE). This technique is based on modification of selected aminoacid residues and has become a standard application of quantitative proteomics. Peptides are labeled with matched sets of fluorescent N-hydroxysuccinimidyl ester cyanines (NHS) that have different excitation-emission wave-lengths (64, 65). Use of thiol-based reagents (maleimide, iodoacetamide) increases specificity and reproducibility. Protein mixtures are label separately with different NHS dyes, combined and resolved on a single 2D gel that is analyzed with different fluorescence excitation wavelengths (66). The differential expression of individual proteins is by this means analyzed and quantified.

Mass Spectrometry

Proteins are generally characterized by mass spectrometry. As ionized molecules are accelerated through an electric field, they are separated, reaching the detector at different times, depending on their mass and charge. This "time-of-flight" (TOF) is specific to a given solute, allowing its identification. Whole proteins need to be first ionized by electrospray ionization (ESI) or matrix-assisted laser desorption/ionization (MALDI). The commonly used "MALDI-TOF" procedure indicates that the MALDI sample preparation is followed by TOF mass analysis, while in the Surface Enhanced Laser Desorption and Ionization (SELDI) procedure, proteins are immobilized on solid supports or customized protein chips.

In the so-called "top-down" strategy, intact proteins are ionized and resolved in the mass analyzer. Alternatively, proteins are pre-digested into smaller peptides before mass analysis; a procedure referred to as "bottom-up". In the latter case, the source protein is identified by its pattern of digestion that creates a "peptide mass fingerprinting" (PMF), or by "*de novo* sequencing", tracking back the protein sequence from the mass sequence data using protein databases. Often, both 'top-down" and ''bottom-up" strategies need to be used to optimize protein identification.

Research Applications

Protein-Protein Interaction

One important question in protein analysis is to identify interactions between proteins, as these are at the core of most intracellular signaling pathways and are critical to the assembly of functional peptides.

Recent developments of the Nat/SDS-PAGE technique permit identification of protein-protein interactions in biological fluids (63). If proteins are extracted from cells or tissues, strategies based on binding to targets linked to solid supports are more advantageously used. These approaches can be further refined using recombinant DNA techniques, in order to construct target protein fragments enabling identification of domains that mediate protein-protein interactions.

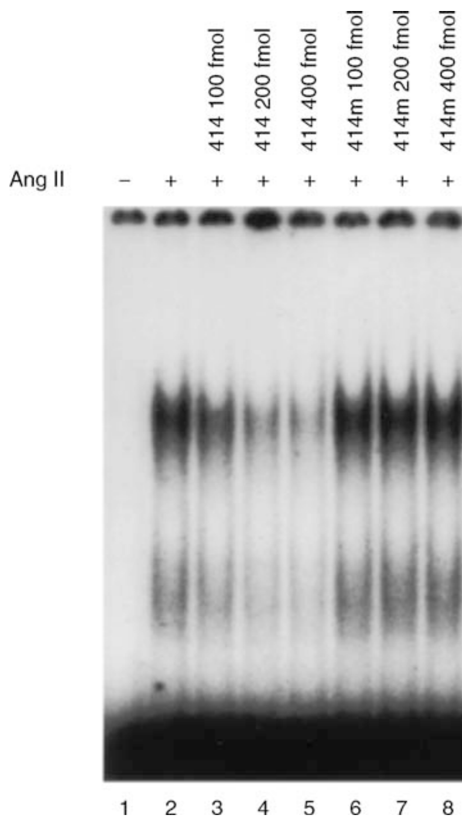
Alternatively, the yeast two-hybrid system is based on the modular structure of gene activation and may be used for the same purpose. In general, the GAL4 transcriptional activator is exploited. This transcription factor has a DNA-binding domain and an activation domain, both of which must be in close association to activate transcription. By DNA recombinant techniques, two protein sequences acting as bait and target are fused with sequences encoding with one of the GAL4 domains. When expressed in yeast cells, the activation domain and the DNA-binding domain are bridged together when the two proteins interact and promote transcription of a reporter gene. This system helps to study interactions between known molecules and to clone new proteins using cDNAs libraries.

Similar prokaryotic systems have been developed based on the modular structure of bacterial RNA polymerases, in which target and bait cDNAs are fused to either the core enzyme or a Φ factor.

Protein-DNA Interaction

Techniques that analyze the regulation of DNA expression by transcriptional factors have helped to identify consensus DNA binding regions and their regulatory proteins. Most of these studies are based on the principle that protein-DNA complexes have high molecular weights and are therefore retarded when resolved by polyacrylamide gel electrophoresis. In addition, proteins interacting with DNA molecules tend to protect the nucleic acid regions to which they bind, from digestion with DNAse, which allows its identification (67). In Fig. 14-1 for example, a nuclear extract of proteins obtained from cells stimulated with angiotensin II was co-incubated with strings of DNA that encode for the promoter of type III collagen. As shown, angiotensin II stimulation promotes synthesis of a protein that binds to the ^{32}P-labeled DNA target, forming macromolecular complexes that are retarded in the gel.

Figure 14-7
Gel retardation assay. Figure demonstrates binding of regulatory nuclear proteins to cis-elements in the COL3A1 promoter after angiotensin II (Ang II) stimulation. Nuclear extracts were obtained from Ang II stimulated cells (Ang II+) and incubated with a 32P-labelled oligonucleotide (414) that contains sequences +3 to +20 of the COL3A1 promoter. Ang II stimulates synthesis of a peptide that binds to the target DNA and retards its migration in the gel (lanes 2). This reaction can be competed with increasing amounts of non-labeled wild-type oligonucleotide (lanes 3–5) but not with a cold mutated analogue sequence (414m) (lanes 6–8). In the absence of Ang II stimulation (AngII −) no DNA-protein complex generating gel retardation is observed (lane 1).



The Building of the Podocyte Protein Map

Unquestionably, the completion of the human genome sequencing has expanded considerably our possibilities to understand and study cell molecular processes. Though over 30,000 genes are encoded in the human DNA however, only a portion is expressed in a give cell. Cell- and tissue-specific cDNA libraries have partially overcome this limitation but do not completely reflect the cell protein repertoire and the levels of expression of individual peptides. Recent advances in protein analysis have now helped to build protein maps. A podocyte protein inventory for example, will allow the study in depth of the signaling pathways regulating cell function in these highly specialized cells, which play a key role in many renal diseases. Definition of the podocyte protein map is in progress and should allow, when completed, to study changes observed under pathological conditions (*Fig. 14-8a*). Currently, podocyte cell lines are being used. Future studies may directly use podocyte expanded from kidney biopsies or from urines of patients with glomerular diseases.

Clinical Applications

Proteomic Analysis of Biological Fluids

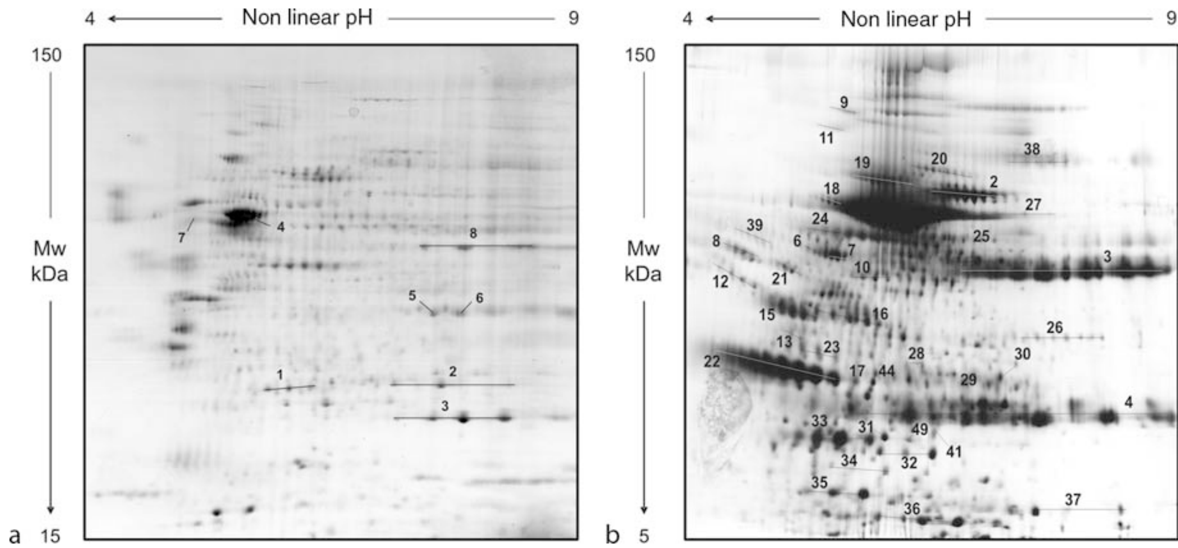
The plasma compartment contains important diagnostic markers and causative molecules of renal diseases. The expression of most plasma proteins cannot be assayed by recombinant DNA technologies, as these are generally not synthesized by circulating cells. Proteomic represents therefore a major tool to study plasma composition. In nephrotic syndrome for example, proteomic has been used to identify glomerular permeability factors and to characterize oxidized protein products. By Nat/SDS PAGE analysis, albumin and other proteins, such as the α1-antitrypsine, have been shown to undergo post-translational modifications that include fragmentation, polymerization, and formation of adducts in nephrotic states (63, 68). Role of these transformed peptide products in the pathophysiology of proteinuria is currently under study.

Likewise, proteomics can also be applied to urine samples (*Fig. 14-8b*). Until recently, urinary proteins were primarily characterized by classic single dimension electrophoresis. Alternatively, the excretion of individual proteins such as albumin, IgG or β2-microglobulin for example, were measured and used as markers for glomerular selectivity or low molecular weight proteinuria.

As the urine proteomic map is nearly completed, researchers are now attempting to use proteomic analysis to create a collection of urinary fingerprints that would allow diagnosis of more specifically renal diseases. Until now, studies in humans have been disappointing and have shown that even distantly related renal diseases can share very similar patterns of proteinuria. The current level of accuracy of urinary biomarkers reliably differentiate

Figure 14-8

Two-dimensional electrophoresis analysis showing a partial normal podocyte proteomic map (a) and a urine proteomic map (b) Identified proteins in Panel A correspond to (1) ubiquitin carboxy terminal; (2) Triosophosphato Isomerase; (3) superoxide dismutase; (4) HSP 60; (5) Glyceraldeyde 3P dehydrogenase; (6) aldose reductase; (7) secernin, (8) alpha enolase. All numbered proteins in Panel B correspond to known proteins that together represent a fingerprint of urinary protein excretion in normal and pathological conditions.



diseases with glomerular and non glomerular involvement, but is inadequate to distinguish between different types of glomerular lesions and to guide their treatment. Experimentally however, candidate disease markers such as haptoglobin in passive heymann nephritis or clusters of proteins in adriamycin nephropathy have been identified (69, 70).

Of notice, urine proteomic analysis also allows to detect factors such as C1qTNF, complement factor Bb or inter-α–trypsin inhibitor chain 4 (spots 50, 23, 52 in *Fig. 14-8b*), which are not detected in plasma because they are readily eliminated in the urine. These proteins also represent potential markers for renal diseases such as primary nephrotic syndrome, IgA nephropathy, or post-transplant proteinuria (71–73).

Peptidome and Degradome

Plasma and urine also contain very small peptides (<5 Kda). Several methodological hurdles still need to be overcome, before applying these analyses to human diseases. Specifically, results obtained by different techniques, such as SELDI and LC-ESI-MS/MS, are often not concordant and lack reproducibility.

Several, if not most of these small urinary peptides, originate from proteolysis of plasma proteins. It is still unclear if their urinary excretion correlates with significant biological events and represent surrogate biomarkers of diseases. Experimentally, a number of them have been shown to have biological activities and most are derived from digestion of urinary albumin.

Few studies have analyzed potentialities of small urinary peptide maps as fingerprints of diseases.

Recently, Decramer et al. have used capillary elctrophoresis followed by tandem mass analysis to characterize the peptide composition of urines obtained from infants born with hydronephrosis secondary to ureteric pelvis junction (74). They could show quantitative changes of a type V pre-procollagen α2 chain fragment that was predictive of the clinical evolution of hydronephrosis and subsequent need for surgery, with a reasonable sensitivity.

Similarly, by combined tandem mass spectrometry, protein chip immunoassay, and SELDI-TOF, O'Riordan et al. have identified two peptides of 4.7 and 4.4 Kda derived from defensin1 and α1antichymotrypsin respectively, and have shown that their ratio correlates with episodes of acute rejection (75). Likewise, they also identified a 4.1 urine peptide that correlates with clinical response in children with nephrotic syndrome (76).

Obviously, most of these findings need to be confirmed with large scale prospective studies. They represent nonetheless one of the forefronts of proteomic research in human diseases and may provide in the near future, less invasive diagnosis and monitoring tools for several renal conditions.

Metabolomics

Similar to proteomics, recent advances in Nuclear Magnetic Resonance (NMR) and Mass Spectrometry (MS) have helped in the search for patterns of recognition of specific conditions by analyzing non-peptidic products of the metabolism. By analogy to "genomics" and "proteomics", these approaches are now referred to as "metabolomics" and their description is beyond the scope of this chapter. The following few examples however suggest that metabolomics may complement proteomics and may provide in the future, new insights into the diagnosis and understanding of renal diseases. In murine models of cis-platinum tubulopathy for example, specific urinary metabolic spectra have been identify that precede the decline in renal function and normalize after treatment (77). Likewise, combined proteomic and metabolomic analysis have identified fingerprints which distinguish between different genetic forms of Fanconi syndrome (78) and specific metabolic profiles have been identified in patients with different glomerular lesions, which may help their diagnosis and monitoring of treatment (79).

References

1. Alberts B, Johnson A, Lewis J et al. Molecular biology of the cell, 5th ed. New York: Garland Publishing Inc, 2008.
2. Lodish H, Berk A, Zipursky SL et al. Molecular Cell Biology, 6th edn. New York, W H Freeman & Co, 2008.
3. Venter JC, Adams MD, Myers EW et al. The sequence of the human genome. *Science* 2001;291:1304–51 (Erratum in: Science. 2001;292:1838).
4. Risch NJ. Searching for genetic determinants in the new millenium. *Nature* 2000;405:847–856.
5. Kininis M, Krauss WL. A global view of transcriptional regulation by nuclear receptors: gene expression, factor localization, and DNA sequence analysis. *Nucl Recept Signal* 2008;6:e005.
6. Grant PA, Schieltz D, Pray-Grant MG et al. A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell* 1998;94:45–53.
7. Tan S, Richmond TJ. Eukaryotic transcription factors. *Curr Opin Struct Biol.* 1998 8:41–48.
8. Weigel NL. Steroid hormone receptors and their regulation by phosphorylation. *Biochem J.* 1996;319:657–667.
9. Moens CB, Selleri L. Hox cofactors in vertebrate development. *Dev Biol.* 2006;291:193–206.
10. Weintraub H, Davis R, Tapscott S et al. The myoD gene family: nodal point during specification of the muscle cell lineage. *Science* 1991;251:761–766.
11. Smart DE, Vincent KJ, Arthur MJ et al. JunD regulates transcription of the tissue inhibitor of metalloproteinases-1 and interleukin-6 genes in activated hepatic stellate cells. *J Biol Chem* 2001;276: 24414–24421.
12. Muller DN, Heissmeyer V, Dechend R et al. Aspirin inhibits NF-kappa B and protects from angiotensin II–induced organ damage. *FASEB J* 2001;15:1822–1824.
13. Kerecuk L, Schreuder MF, Woolf AS. Renal tract malformations: perspectives for nephrologists. *Nat Clin Pract Nephrol 2008*;4:312–35.
14. Devarajan P, Mishra J, Supavekin S et al. Gene expression in early ischemic renal injury: clues towards pathogenesis, biomarker discovery, and novel therapeutics. *Mol Genet Metab* 2003;80:365–376.
15. Calarco JA, Saltzman AL, Ip JY, Blencowe BJ. Technologies for the global discovery and analysis of alternative splicing. *Adv Exp Med Biol.* 2007;623:64–84.
16. Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell* 2006;126:37–47.
17. Modrek B, Lee C. A genomic view of alternative splicing. *Nat Genet* 2002;30:13–19.
18. Klamt B, Koziell A, Poulat F et al. Frasier syndrome iscaused by defective alternative splicing of WT1 leading to an altered ratio of WT1 + /–KTS splice isoforms. *Hum Mol Genet* 1998;7:709–714.
19. Anderson JT. RNA turnover: unexpected consequences of being tailed. *Curr Biol* 2005;15:R635–R638.
20. Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* 2007;8:113–126.
21. Cleveland DW. Gene regulation through messenger RNA stability. *Curr Opin Cell Biol* 1989;1:1148–1153.
22. Holmes CL, Landry DW, Granton JT. Science review: Vasopressin and the cardiovascular system part 1 – Receptor physiology. *Crit Care* 2003;7:427–434.
23. Ben-Neriah Y. Regulatory functions of ubiquitination in the immune system. *Nat Immunol* 2002;3:20–26.
24. Nathans D, Smith HO. Restriction endonucleases in the analysis and restructuring of DNA molecules. *Annu Rev Biochem* 1975;44: 273–293.
25. Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase chain reaction. *Methods Enzymol* 1987;155:335–350.
26. Prediger EA. Quantitating mRNAs with relative and competitive RT-PCR. *Methods Mol Biol* 2001;160:49–63.
27. Wittwer CT, Herrmann MG, Gundry CN et al. Real-time multiplex PCR assays. *Methods* 2001;25:430–442.
28. Von Schnakenburg C, Strehlau J, Ehrich JH et al. Quantitative-gene expression of TGF-beta1, IL-10, TNF-alpha and Fas Ligand in renal cortex and medulla. *Nephrol Dial Transpl* 2002;17: 573–579.
29. Böttinger EP, Ju W, Zavadil Applications for microarrays in renal biology and medicine. *J Exp Nephrol* 2002;10:93–101.
30. Duggan DJ, Bittner M, Chen Y et al. Expression profilingusing cDNA microarrays. *Nat Genet* 1999;21(Suppl):10–14.
31. Royce TE, Rozowsky JS, Bertone P et al. Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping. *Trends Genet* 2005;21:466–475.
32. Hom G, Graham RR, Modrek B et al. Association of Systemic Lupus Erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 358:900–909.
33. Harbers M, Carninci P. Tag-based approaches for transcriptome research and genome annotation. *Nat Methods* 2005;2:495–502.

34. Koenen M, Griesser H-W, Muller-Hill B. Immunological detection of chimeric α-galactosidases expressed by plasmid vectors. In DNA Cloning Techniques: A Practical approach. Glover DM (ed.). Oxford, IRL Press 1984, pp. 89–100.
35. Yayon A, Aviezer D, Saran M et al. Isolation of peptides that inhibit binding of basic fibroblast growth factor to its receptor from a random phage epitope library. *Proc NatlAcad Sci USA* 1993;90: 10643–10647.
36. Brown EM, Gamba G, Riccardi D et al. Cloning and characterizationof an extracellular Ca(2+)-sensing receptor frombovine parathyroid. *Nature* 1993;366:575–580.
37. Henikoff S, Greene EA, Pietrokovski S et al. Gene families:the taxonomy of protein paralogs and chimeras. *Science* 1997;278:609–614.
38. Obinata M. The immortalized cell lines with differentiation potentials: their establishment and possible application. *Cancer Sci* 2007;98:275–283.
39. Kalatzis V, Nevo N, Cherqui S et al. Molecular pathogenesis of cystinosis: effect of CTNS mutations on the transport activity and subcellular localization of cystinosin. *Hum Mol Genet* 2004;13: 1361–1371.
40. Anderson MP, Gregory RJ, Thompson S et al. Demonstration that CFTR is a chloride channel by alteration of itsanion selectivity. *Science* 1991;253:202–207.
41. Orlowski J, Grinstein S. Na + /H + exchangers of mammalian cells. *J Biol Chem* 1997;272:22373–22376.
42. Gamba G, Miyanoshita A, Lombardi M et al. Molecular cloning, primary structure and characterization of two members of the mammalian electroneutral sodium-(potassium)-chloride cotransporter family expressed in kidney. *J Biol Chem* 1994;269: 17713–17722.
43. Ho K, Nichols CG, Lederer WJ et al. Cloning and expression of an inwardly rectifying ATP-regulated potassium channel. *Nature* 1993;362:31–38.
44. Chrispeels MJ, Agre P. Aquaporins: water channel proteins of plant and animal cells. *Trends Biochem Sci* 1994;19:421–425.
45. Canessa CM, Schild L, Buell G et al. Amiloride-sensitive epithelial Na + channel is made of three homologous subunits. *Nature* (Lond) 1994;367:463–467.
46. Deen PMT, Verdijk MAJ, Knoers NVAM et al. Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. *Science* 1994;264:92–95.
47. Simon RP, Lifton RP. The molecular basis of inherited hypokalemic alkalosis: Bartter's and Gitelman's syndromes. *Am J Physiol* 1996;271: F961–F969.
48. Karolyil L, Konrad M, Kockerling A et al. Mutations in the gene encoding the inwardly-rectifying renal potassium channel, ROMK, cause the antenatal variant of Bartter syndrome: evidence for genetic heterogeneity. *Hum Mol Genet* 1997;6:17–26.
49. Lemmink HH, van den Heuvell LPWJ, van Dijk HA et al. Linkage of Gitelman syndrome to the human thiazide-sensitive sodium-chloride cotransporter gene with identification of mutations in three Dutch families. *Pediatr Nephrol* 1996;10:403–407.
50. Hannson JH, Schild L, Lu Y et al. A de novo missense mutation of the beta subunit of the epithelial sodium channel causes hypertension and Liddle's syndrome, identifying a proline-rich segment critical for regulation of channel activity. *Proc Natl Acad Sci USA* 1995;92: 11495–11499.
51. Behm-Ansmant I, Rehwinkel J, Izaurralde E. MicroRNAs silence gene expression by repressing protein expression and/or by promoting mRNA decay. *Cold Spring Harb Symp Quant Biol* 2006;71: 523–530.
52. Miller LK. Baculoviruses as gene expression vectors. *Annu Rev Microbiol* 1988;42:177–199.
53. Suter B, Auerbach D, Stagljar I. Yeast-based functional genomics and proteomics technologies: the first 15 years and beyond. *Biotechniques* 2006;40:625–44.
54. Hanahan D. Transgenic mice as probes into complex systems. *Science* 1989;246:1265–1274.
55. Musaro A, Rosenthal N. Transgenic mouse models of muscle aging. *Exp Gerontol* 1999;34:147–156.
56. Kohan DE. Progress in gene targeting: Using mutant mice to study renal function and disease. *Kidney Int* 2008;74:427–47.
57. Sasaki S, Nishihira J, Ishibashi T et al. Transgene of MIF induces podocyte injury and progressive mesangial sclerosis in the mouse kidney. *Kidney Int* 2004;65:469–481.
58. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975;250:4007–4021.
59. Anderson L, Anderson NG. High resolution two-dimensional electrophoresis of human plasma proteins. *Proc Natl Acad Sci USA* 1977;74:5421–5425.
60. Bruschi M, Musante L, Candiano G et al. Soft immobilized pH gradient gels in proteome analysis: a follow-up. *Proteomics* 2003;3: 821–825.
61. Manabe T, Jin Y. Analysis of protein/polypeptide interactions in human plasma using nondenaturing micro-2-DE followed by 3-D SDS-PAGE and MS. *Electrophoresis* 2007;28:2065–2079.
62. Manabe T, Jin Y. Tani O, Assignment of human plasma polypeptides on a nondenaturing 2-D gel using MALDI-MS and PMF and comparisons with the results of intact protein mapping. *Electrophoresis* 2007;28:843–863.
63. Candiano G, Santucci L, Petretto A et al. High-resolution 2-DE for resolving proteins, protein adducts and complexes in plasma. *Electrophoresis* 2008, 29, 682–694.
64. Gharbi S, Gaffney P, Yang A et al. Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol Cell Proteomics* 2002, 1, 91–98.
65. Alban A, David SO, Bjorkesten L et al. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* 2003;3:36–44.
66. Knowles MR, Cervino S, Skynner HA et al. Multiplex proteomic analysis by two-dimensional differential in-gel electrophoresis. *Proteomics* 2003;3:1162–1171.
67. Rosenthal N. Molecular medicine. Recognizing DNA. *N Engl J Med* 1995;333:925–927.
68. Candiano G, Musante L, Bruschi et al. Repetitive fragmentation products of albumin and alpha1-antitrypsin in glomerular diseases associated with nephrotic syndrome. *J Am Soc Nephrol* 2006;17: 3139–3148.
69. Ngai HH, Sit WH, Jiang PP, ThongboonkerdV et al. Markedly increased urinary preprohaptoglobin and haptoglobin in passive heymann nephritis: a differential proteomics approach. *J Proteome Res* 2007;6:3313–3320.
70. Shui HA, Huang TH, Ka SM et al. Urinary proteome and potential biomarkers associated with serial pathogenesis steps of focal segmental glomerulosclerosis. *Nephrol Dial Transpl* 2008;23:176–185.
71. Varghese SA, Powell TB, Budisavljevic MN et al. Urine biomarkers predict the cause of glomerular disease. *J Am Soc Nephrol* 2007;18:913–922.
72. Rocchetti MT, Centra M, Papale M et al. Urine protein profile of IgA nephropathy patients may predict the response to ACE-inhibitor therapy. *Proteomics* 2008;8:206–216.

73. Candiano G, Bruschi M, Petretto A et al. Urinary Proteins, Protein Fragments and Peptides in Nephrotic Syndrome. Clusters, Specficity and Mechanisms. *Proteomics Clin Appl* 2008;2:956–963.
74. Decramer S, Wittke S, Mischak H et al. Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis. *Nat Med* 2006;12:398–400.
75. O'Riordan E, Orlova TN, Podust et al. Characterization of urinary peptide biomarkers of acute rejection in renal allografts. *Am J Transpl* 2007;7:930–940.
76. Woroniecki RP, Orlova TN, Mendelev N et al. Urinary proteome of steroid-sensitive and steroid-resistant idiopathic nephrotic syndrome of childhood. *Am J Nephrol* 2006;26:258–267.
77. Portilla D, Schnackenberg L, Beger RD. Metabolomics as an extension of proteomic analysis: study of acute kidney injury. *Semin Nephrol* 2007;27:609–620.
78. Vilasi A, Cutillas PR, Maher AD et al. Combined proteomic and metabonomic studies in three genetic forms of the renal Fanconi sindrome. *Am J Physiol Renal Physiol* 2007;293:F456–F467.
79. Psihogios NG, Kalaitzidis RG, Dimou S et al. Evaluation of tubulointerstitial lesions' severity in patients with glomerulonephritides: an NMR-based metabonomic study. *J Proteome Res* 2007;6:3760–3770.