

## A GLOSSARY OF TERMS COMMONLY USED IN MOLECULAR BIOLOGY\*

Stanley M. Lemon, M.D. and Alan G. Barbour, M.D.  
Department of Medicine, The University of North Carolina at Chapel Hill  
Chapel Hill, NC 27599-7030  
Departments of Microbiology and Medicine, University of Texas Health Science Center  
San Antonio, TX 78284-7758  
(Copyright, S.M. Lemon and A.G. Barbour, 1993)

\*Small Additions made March 2011-CAM

**AGAROSE GEL ELECTROPHORESIS** - A method for separating nucleic acids (DNA or RNA) within a gel made of agarose in a suitable buffer under the influence of an electrical field. Suitable for separation of large fragments of nucleic acid, separation is based primarily upon the size of the nucleic acid.

**ALLELE** - One of several alternate forms of a gene occupying a given locus on a chromosome or plasmid.

**AMINO ACIDS** - The 20 basic building blocks of proteins, consisting of the basic formula  $\text{NH}_2\text{-CHR-COOH}$ , where "R" is the side chain which defines the amino acid:

Nonpolar side chains (hydrophobic)

G Gly Glycine  
A Ala Alanine  
V Val Valine  
I Ile Isoleucine  
L Leu Leucine  
F Phe Phenylalanine  
P Pro Proline  
M Met Methionine  
W Trp Tryptophan  
C Cys Cysteine

Noncharged polar side chains (hydrophilic)

S Ser Serine  
T Thr Threonine  
Y Tyr Tyrosine  
N Asn Asparagine  
Q Gln Glutamine

Acidic side chains (very polar, hydrophilic)

D Asp Aspartic Acid  
E Glu Glutamic Acid

Basic side chains (very polar, hydrophilic)

K Lys Lysine  
R Arg Arginine  
H His Histidine

**AMINO TERMINUS** - Refers to the  $\text{NH}_2$  end of a *peptide* chain (by custom drawn at the left of a protein sequence)

**AMPLIFICATION** - Refers to the production of additional copies of a chromosomal sequence, found either as intrachromosomal or extrachromosomal DNA. Also refers to the in vitro process in the *polymerase chain reaction*.

**AMPLIMER** - Region of DNA sequence which is amplified during a *PCR* reaction and which is defined by a pair of *PCR primers* (these *primer* pairs are sometimes called amplimers).

**ANCHOR SEQUENCE** - A hydrophobic amino acid sequence which fixes a segment of a newly synthesized, *translocating* protein within the lipid bilayer membrane of the endoplasmic reticulum.

**ANNEAL** - See HYBRIDIZATION.

**ANTISENSE STRAND (OR PRIMER)** - Refers to the RNA or DNA strand of a *duplex* molecule which is *complementary* to that encoding a *polypeptide*. More specifically, the DNA strand which serves as *template* for the synthesis of RNA and which is complementary to it. "Antisense *oligonucleotides*" hybridize to *mRNA*, and are used to prime *cDNA* synthesis.

**ASSEMBLED EPITOPE** - See CONFORMATIONAL EPITOPE.

**AUTORADIOGRAPHY** - A process to detect radioactively labeled molecules (which usually have been separated in an *SDS-PAGE* or *agarose gel*) based on their ability to create an image on photographic or X-ray film. This process does not result in a linear relationship between the intensity of the signal and the amount of radioactivity unless special steps are taken. There is now increasing use of phosphorimagers and other modern devices to detect and quantitate radioactive molecules which have been separated in gels.

**AVIDIN** - A glycoprotein which binds to *biotin* with very high affinity ( $K_d = 10^{-15}$ ).

**BACK MUTATION** - Reverse the effect of a point or frame-shift mutation that had altered a gene; thus it restores the wild-type phenotype (see REVERTANT).

**BACTERIOPHAGE** - A virus that infects bacteria; often simply called a phage. The phages which are most often used in molecular biology are the *E. coli* viruses lambda, M13 and T7.

**BASE** - The *purine* or *pyrimidine* component of a *nucleotide*; often used to refer to a *nucleotide* residue within a nucleic acid chain.

**BASE PAIR** - One pair of *complementary nucleotides* within a *duplex* strand of a nucleic acid. Under Watson-Crick rules, these pairs consist of one *pyrimidine* and one *purine*: i.e., C-G, A-T (DNA) or A-U (RNA). However, "noncanonical" base pairs (e.g., G-U) are common in RNA *secondary structure*.

**BIOTIN** - A coenzyme which is essential for carboxylation reactions (see AVIDIN).

**BLUNT END** - A terminus of a *duplex* DNA molecule which ends precisely at a base pair, with no *overhang* (unpaired nucleotide) in either strand. Some but not all *restriction endonucleases* leave blunt ends after cleaving DNA. Blunt-ended DNA can be *ligated* nonspecifically to other blunt-ended DNA molecules (compare with STICKY END).

5'-->3'  
NNNCCC GGGNNN *Sma*I cut, no overhang  
NNNGGG CCCNNN  
3'<--5'

**bp** - "base pair"

**BOX** - Refers to a short nucleic acid *consensus sequence* or *motif* that is universal within kingdoms of organisms. Examples of DNA boxes are the Pribow box (TATAAT) for *RNA polymerase*, the Hogness box (TATA) that has a similar function in eukaryotic organisms, and the homeo box. RNA boxes have

also been described, such as Pilipenko's Box-A motif that may be involved in ribosome binding in some viral RNAs.

**C TERMINUS** - See CARBOXYL TERMINUS.

**CARBOXYL TERMINUS** - Refers to the COOH end of a *peptide* chain (by custom drawn at the right of a protein sequence)

**cDNA** - Complementary DNA. A DNA molecule which was originally copied from an RNA molecule by *reverse transcription*. The term "cDNA" is commonly used to describe double-stranded DNA which originated from a single-stranded RNA molecule, even though only one strand of the DNA is truly complementary to the RNA.

**cDNA LIBRARY** - A collection of *cDNA* fragments, each of which has been cloned into a separate *vector* molecule.

**CAP** - A 7-methyl guanosine residue linked 5' to 5' through a triphosphate bridge to the 5' end of eukaryotic *mRNAs*; facilitates initiation of *translation*.

**CELL**- The lowest denomination of life thought to be possible. Most organisms consist of more than one cell which becomes specialized into particular functions (usually expressing certain identifiable proteins associated with a specific function) to enable the whole organism to function properly. Cells contain DNA and many other elements to enable the cell to survive to divide and to function.

**CHAIN TERMINATOR** - See DIDEOXYNUCLEOTIDE.

**CHAPERONE PROTEINS** - A series of proteins present in the endoplasmic reticulum which guide the proper folding of secreted proteins through a complex series of binding and release reactions.

**CHROMOSOMES** - The self-replicating genetic structure of cells containing the cellular DNA. Humans have 23 pairs of chromosomes.

**CHROMOSOME WALKING** - The sequential isolation of *clones* carrying overlapping sequences of DNA which span large regions of a chromosome. Overlapping regions of clones can be identified by *hybridization*.

**CLONE** - Describes a large number of cells, viruses, or molecules which are identical and which are derived from a single ancestral cell, virus or molecule. The term can be used to refer to the process of isolating single cells or viruses and letting them proliferate (as in a *hybridoma* clone, which is a "biological clone"), or the process of isolating and replicating a piece of DNA by recombinant DNA techniques ("molecular clone"). The use of the word as a verb is acceptable for the former meaning, but not necessarily the latter meaning.

**CIS** - As used in molecular biology, an interaction between two sites which are located within the same molecule. However, a *cis*-acting protein can either be one which acts only on the molecule of DNA from which it was expressed, or a protein which acts on itself (e.g., self-proteolysis).

**CISTRON** - A nucleic acid segment corresponding to a polypeptide chain, including the relevant *translational start (initiation)* and *stop (termination) codons*.

**CODON** - A nucleotide triplet (sequence of three nucleotides) which specifies a specific *amino acid*, or a *translational* start or stop.

**CODON BIAS** - The tendency for an organism or virus to use certain codons more than others to encode a particular amino acid. An important determinant of codon bias is the guanosine-cytosine (GC)

content of the genome. An organism that has a relatively low G+C content of 30% will be less likely to have a G or C at the third position of a codon (wobble position) than a A or T to specify an amino acid that can be represented by more than one codon.

**COLONY FORMING UNITS** – a way to measure the number of viable bacteria in a suspension by diluting the suspension then plating out dilutions onto nutrient agar plates and incubating the plates such that one initial bacteria multiplies during the incubation to form one colony of cells. Usually expressed as cfu/mL or some other measure of volume.

**COMPETENT** - Bacterial cells which are capable of accepting foreign extra-chromosomal DNA. There are a variety of processes by which cells may be made competent.

**COMPLEMENTARY** - See BASE PAIR.

**CONFORMATIONAL EPITOPE** - An epitope which is dependent upon folding of a protein; amino acid residues present in the antibody binding site are often located at sites in the primary sequence of the protein which are at some distance from each other. The vast majority of B-cell (antibody binding) epitopes are conformational.

**CONJUGATION** – a process in bacteria where plasmids are mobilized and transferred between bacterial cells. This usually involves the expression of specific factors (FR) in at least one bacteria. Plasmids as well as bacterial chromosomes can transferred this way.

**CONSENSUS SEQUENCE** - A linear series of nucleotides, commonly with gaps and some degeneracy, that define common features of homologous sequences or recognition sites for proteins that act on or bind to nucleic acids.

**CONSERVATIVE SUBSTITUTION** - A nucleotide mutation which alters the amino acid sequence of the protein, but which causes the substitution of one amino acid with another which has a side chain with similar charge/polarity characteristics (see AMINO ACID). The size of the side chain may also be an important consideration. Conservative mutations are generally considered unlikely to profoundly alter the structure or function of a protein, but there are many exceptions (see NONCONSERVATIVE SUBSTITUTION).

**CONSERVED** - Similar in structure or function.

**CONTIG** - A series of two or more individual DNA sequence determinations that overlap. In a sequencing project the contigs get larger and larger until the gaps between the contigs are filled in.

**COSMID** - A genetically-engineered *plasmid* containing bacteriophage lambda packaging signals and potentially very large pieces of inserted foreign DNA (up to 50 kb) which can be replicated in bacterial cells. Cosmid cloning allows for isolation of DNA fragments which are larger than those which can be cloned in conventional plasmids.

**DATABASE SEARCH** - Once an open *reading frame* or a partial amino acid sequence has been determined, the investigator compares the sequence with others in the databases using a computer and a search algorithm. This is usually done in a protein database such as PIR or Swiss-Prot. Nucleic acid sequences are in GenBank and EMBL databases. The search algorithms most commonly used are BLAST and FASTA.

**DEGENERACY** - Refers to the fact that multiple different *codons* in *mRNA* can specify the same *amino acid* in an encoded protein.

**DENATURATION** - With respect to nucleic acids, refers to the conversion from double-stranded to the single-stranded state, often achieved by heating or alkaline conditions. This is also called "melting"

DNA. With respect to proteins, refers to the disruption of *tertiary* and *secondary* structure, often achieved by heat, detergents, chaotropes, and sulfhydryl-reducing agents.

**DENATURING GEL** - An *agarose* or *acrylamide gel* run under conditions which destroy *secondary* or *tertiary* protein or RNA structure. For protein, this usually means the inclusion of 2-ME (which reduces disulfide bonds between cysteine residues) and SDS and/or urea in an acrylamide gel. For RNA, this usually means the inclusion of formaldehyde or glyoxal to destroy higher ordered RNA structures. In DNA sequencing gels, urea is included to denature dsDNA to ssDNA strands. In denaturing gels, macromolecules tend to be separated on the basis of size and (to some extent) charge, while shape and oligomerization of molecules are not important. Contrast with NATIVE GEL.

**DEOXYRIBONUCLEASE (DNase)** - An enzyme which specifically catalyzes the hydrolysis of DNA.

**DEOXYRIBONUCLEOTIDE** - *nucleotides* which are the building blocks of DNA and which lack the 2' hydroxyl moiety present in the ribonucleotides of RNA.

**DIDEOXYRIBONUCLEOTIDE** - A *nucleotide* which lacks both 3' and 2' hydroxyl groups. Such dideoxynucleotides can be added to a growing nucleic acid chain, but do not then present a 3' -OH group which can support further propagation of the nucleic acid chain. Thus such compounds are also called "chain terminators", and are useful in DNA and RNA sequencing reactions (see DEOXYRIBONUCLEOTIDE).

**DIDEOXY SEQUENCING** - Enzymatic determination of DNA or RNA sequence by the method of Sanger and colleagues, based on the incorporation of chain terminating *dideoxynucleotides* in a growing nucleic acid strand copied by *DNA polymerase* or *reverse transcriptase* from a DNA or RNA *template*. Separate reactions include dideoxynucleotides containing A, C, G, or T bases. The reaction products represent a collection of new, labeled DNA strands of varying lengths, all terminating with a dideoxynucleotide at the 3' end (at the site of a complementary base in the template nucleic acid), and are separated in a polyacrylamide/urea gel to generate a sequence "ladder". This method is more commonly used than "Maxam-Gilbert" (chemical) sequencing.

**DIRECT REPEATS** - Identical or related sequences present in two or more copies in the same orientation in the same molecule of DNA; they are not necessarily adjacent.

**DNA LIGASE** - An enzyme (usually from the T4 bacteriophage) which catalyzes formation of a phosphodiester bond between two adjacent bases from double-stranded DNA fragments. RNA ligases also exist, but are rarely used in molecular biology.

**DNA POLYMERASE** - A polymerase which synthesizes DNA (see POLYMERASE).

**DNase** - see DEOXYRIBONUCLEASE.

**DOT BLOT** - DNA or RNA is simply spotted onto nitrocellulose or nylon membranes, denatured and hybridized with a *probe*. Unlike *Southern* or *northern blots*, there is no separation of the target DNA or RNA by electrophoresis (size), and thus potentially much less specificity.

**DOWNSTREAM** - Identifies sequences proceeding farther in the direction of *expression*; for example, the coding region is downstream from the *initiation codon*, toward the 3' end of an mRNA molecule. Sometimes used to refer to a position within a protein sequence, in which case downstream is toward the *carboxyl* end which is synthesized after the *amino* end during translation.

**ds** - "double-stranded"

**DUPLEX** - A nucleic acid molecule in which two strands are *base paired* with each other.

**ELECTROPORATION** - A method for introducing foreign nucleic acid into bacterial or eukaryotic cells that uses a brief, high voltage DC charge which renders the cells permeable to the nucleic acid. Also useful for introducing synthetic peptides into eucaryotic cells.

**END LABELING** - The technique of adding a radioactively labeled group to one end (5' or 3' end) of a DNA strand.

**ENDONUCLEASE** - Cleaves bonds within a nucleic acid chain; they may be specific for RNA or for single-stranded or double-stranded DNA. A restriction enzyme is a type of endonuclease.

**ENHANCER** - A eukaryotic *transcriptional* control element which is a DNA sequence which acts at some distance to enhance the activity of a specific *promoter* sequence. Unlike promoter sequences, the position and orientation of the enhancer sequence is generally not important to its activity.

**EPITOPE** - As related to protein antigens, B-cell epitopes consist of the amino acid residues of a protein molecule which interact directly through noncovalent bonds with the amino acid residues of a particular antibody molecule (complementarity determining region). The average epitope probably involves about 15-20 contact amino acid residues, but one or two of these may be critical to the epitope's specificity and the avidity of the antibody-antigen reaction. B-cell epitopes may be either *linear* or *conformational* in nature. T-cell epitopes represent the small, processed peptides which bind to MHC class I and II molecules on the surface of T cells.

**ETHIDIUM BROMIDE** - Intercalates within the structure of nucleic acids in such a way that they fluoresce under UV light. Ethidium bromide staining is commonly used to visualize RNA or DNA in agarose gels placed on UV light boxes. Proper precautions are required, because the ethidium bromide is highly mutagenic and the UV light damaging to the eyes. Ethidium bromide is also included in cesium chloride gradients during ultracentrifugation, to separate *supercoiled* circular DNA from linear and *relaxed* circular DNA.

**EVOLUTIONARY CLOCK** - Defined by the rate at which mutations accumulate within a given gene.

**EXON** - The portion of a gene that is actually translated into protein (see INTRON, SPLICING).

**EXONUCLEASE** - An enzyme which hydrolyzes DNA beginning at one end of a strand, releasing nucleotides one at a time (thus, there are 3' or 5' exonucleases)

**EXPRESSION** - Usually used to refer to the entire process of producing a protein from a gene, which includes *transcription*, *translation*, *post-translational modification* and possibly transport reactions.

**EXPRESSION VECTOR** - A plasmid or phage designed for production of a polypeptide from inserted foreign DNA under specific controls. Often an *inducer* is used. The vector always provides a promoter and often the *transcriptional start site*, *ribosomal binding sequence*, and *initiation codon*. In some cases the product is a *fusion protein*.

**FOOTPRINTING** - A technique for identifying the site on a DNA (or RNA) molecule which is bound by some protein by virtue of the protection afforded *phosphodiester bonds* in this region against attack by *nuclease* or nucleolytic compounds.

**FRAMESHIFT MUTATION** - A mutation (deletion or insertion, never a simple substitution) of one or more *nucleotides* but never a multiple of 3 nucleotides, which shortens or lengthens a trinucleotide sequence representing a *codon*; the result is a shift from one *reading frame* to another reading frame. The amino acid sequence of the protein downstream of the mutation is completely altered, and may even be much shorter or longer due to a change in the location of the first *termination (stop) codon*:

Asn Tyr Thr Asn Leu Gly His Wild-type polypeptide

AAU UAC ACA AAU UUA GGG CAU mRNA  
Asn Thr Gln Ile STOP Mutant polypeptide

|  
Deletion of A from mRNA creates frame-shift mutant

**FUSION PROTEIN** - A product of recombinant DNA in which the foreign gene product is juxtaposed ("fused") to either the *carboxyl-terminal* or *amino-terminal* portion of a polypeptide encoded by the vector itself. Use of fusion proteins often facilitates expression of otherwise lethal products and the purification of recombinant proteins.

**GEL SHIFT** - A method by which the interaction of a nucleic acid (DNA or RNA) with a protein is detected. The mobility of the nucleic acid is monitored in an agarose gel in the presence and absence of the protein: if the protein binds to the nucleic acid, the complex migrates more slowly in the gel (hence "gel shift"). A "supershift" allows determination of the specific protein, by virtue of a second shift in mobility that accompanies binding of a specific antibody to the nucleic acid-protein complex.

**GENE** - Generally speaking, the *genomic* nucleotide sequence that codes for a particular polypeptide chain, including relevant *transcriptional* control sequences and *introns* (if a eukaryote). However, the term is often loosely used to refer to only the relevant coding sequence.

**GENE CONVERSION** - The alteration of all or part of a gene by a homologous donor DNA that is itself not altered in the process.

**GENOME** - The complete set of genetic information defining a particular animal, plant, organism or virus.

**GENOMIC LIBRARY** - A DNA library which contains DNA fragments hopefully representing each region of the genome of an organism, virus, etc, cloned into individual vector molecules for subsequent selection and amplification. The DNA fragments are usually very small in size compared with the genome. Such libraries are *cDNA libraries* when prepared from RNA viruses.

**GENOTYPE** - The genetic constitution of an organism; determined by its nucleic acid sequence. As applied to viruses, the term implies a group of evolutionarily related viruses possessing a defined degree of nucleotide sequence relatedness.

**GLYCOPROTEIN** - A *glycosylated* protein.

**GLYCOSYLATION** - The covalent addition of sugar moieties to N or O atoms present in the side chains of certain amino acids of certain proteins, generally occurring within the Golgi apparatus during secretion of a protein. Glycosylation sites are only partially predictable by current computer searches for relevant motifs in protein sequence. Glycosylation may have profound but very unpredictable effects on the folding, stability, and antigenicity of secreted proteins. Glycosylation is a property of eukaryotic cells, and differs among different cell types (i.e., it may be very different in yeast or insect cells used for protein expression, when compared with Chinese hamster ovary (CHO) cells).

**GOLGI APPARATUS** - A membranous, vesicular structure which is in continuity with the *endoplasmic reticulum* of eukaryotic cells and generally in close proximity to the nucleus, the Golgi plays an important role in the posttranslational *processing* and transport of secreted proteins.

**HAIRPIN** - A helical (duplex) region formed by base pairing between adjacent (inverted) complementary sequences within a single strand of RNA or DNA.

**HETERODUPLEX DNA** - Generated by base pairing between complementary single strands derived from different parental *duplex* molecules; heteroduplex DNA molecules occur during genetic *recombination* in vivo and during *hybridization* of different but related DNA strands in vitro. Since the

sequences of the two strands in a heteroduplex differ, the molecule is not perfectly base-paired; the *melting* temperature of a heteroduplex DNA is dependent upon the number of mismatched base pairs.

**HIV – Human Immunodeficiency Virus**, the causative agent of AIDS, is a lentivirus, related to Retroviruses.

**HOMOLOGOUS RECOMBINATION** - The exchange of sequence between two related but different DNA (or RNA) molecules, with the result that a new "chimeric" molecule is created. Several mechanisms may result in recombination, but an essential requirement is the existence of a region of *homology* in the recombination partners. In DNA recombination, breakage of single strands of DNA in the two recombination partners is followed by joining of strands present in opposing molecules, and may involve specific enzymes. Recombination of RNA molecules may occur by other mechanisms.

**HOMOLOGY** - Indicates similarity between two different *nucleotide* or *amino acid* sequences, often with potential evolutionary significance. It is probably better to use more quantitative and descriptive terms such as nucleotide "identity" or, in the case of proteins, amino acid "identity" or "relatedness" (the latter refers to the presence of amino acids residues with similar polarity/charge characteristics at the same position within a protein).

**HYBRIDIZATION** - The process of *base pairing* leading to formation of *duplex* RNA or DNA or RNA-DNA molecules.

**HYBRIDOMA** - A *clone* of plasmacytoma cells which secrete a monoclonal antibody; usually produced by fusion of peripheral or splenic plasma cells taken from an immunized mouse with an immortalized murine plasmacytoma cell line (fusion partner), followed by *cloning* and *selection* of appropriate antibody-producing cells.

**HYDROPHILICITY PLOT** - A computer plot which examines the relative summed hydrophobicity/hydrophilicity of adjacent *amino acid* sidechains (usually within a moving window of about 6 amino acid residues) along the primary sequence of a polypeptide chain. Values for the contribution of sidechains of each the 20 common amino acids to hydrophobicity/hydrophilicity have been developed by Hopp & Woods, and Kyte & Doolittle, and these plots are often named after these workers. Generally, hydrophobic regions of proteins are considered likely to be in the interior of the native protein, while hydrophilic domains are likely to be exposed on the surface and thus possibly antigenic sites (epitopes). At best, these are crude predictions.

**IMMUNOBLOT** - See WESTERN BLOT.

**IMMUNOPRECIPITATION** - A process whereby a particular protein of interest is isolated by the addition of a specific antibody, followed by centrifugation to pellet the resulting immune complexes. Often, staphylococcal proteins A or G, bound to sepharose or some other type of macroscopic particle, is added to the reaction mix to increase the size and ease collection of the complexes. Usually, the precipitated protein is subsequently examined by *SDS-PAGE*.

**INDUCER** - A small molecule, such as IPTG, that triggers gene transcription by binding to a regulator protein, such as LacZ.

**INITIATION CODON** - The *codon* at which translation of a polypeptide chain is initiated. This is usually the first AUG triplet in the *mRNA* molecule from the 5' end, where the ribosome binds to the cap and begins to scan in a 3' direction. However, the surrounding sequence context is important and may lead to the first AUG being bypassed by the scanning ribosome in favor of an alternative, downstream AUG. Also called a "start codon". Occasionally other codons may serve as initiation codons, e.g. UUG.

**INSERT** - Foreign DNA placed within a vector molecule.



**INSERTION SEQUENCE** - A small bacterial transposon carrying only the genetic functions involved in *transposition*. There are usually *inverted repeats* at the ends of the insertion sequence.

**INTRON** - Intervening sequences in eukaryotic genes which do not encode protein but which are *transcribed* into RNA. Removed from *pre-mRNA* during nuclear *splicing* reactions.

**INVERTED REPEATS** - Two copies of the same or related sequence of DNA repeated in opposite orientation on the same molecule (contrast with DIRECT REPEATS). Adjacent inverted repeats constitute a *palindrome*.

**IN VITRO TRANSLATION** - See RETICULOCYTE LYSATE.

**kb** - 'kilobase'

**KILOBASE** - Unit of 1000 nucleotide bases, either RNA or DNA.

**KINASE** - See PHOSPHORYLATION.

**KLENOW FRAGMENT** - The large fragment of E. coli DNA polymerase I which lacks 5' → 3' exonuclease activity. Very useful for sequencing reactions, which proceed in a 5' → 3' fashion (addition of nucleotides to templated free 3' ends of primers).

**KNOCK-OUT** - The excision or inactivation of a gene within an intact organism or even animal (e.g., "knock-out mice"), usually carried out by a method involving *homologous recombination*.

**LIBRARY** - A set of cloned fragments together representing with some degree of redundancy the entire genetic complement of an organism (see cDNA LIBRARY, GENOMIC LIBRARY).

**LIGASE** - See DNA LIGASE.

**LIGATION** - See DNA LIGASE.

**LINEAR EPITOPE** - An epitope formed by a series of amino acids which are adjacent to each other within the primary structure of the protein. Such epitopes can be successfully modelled by synthetic peptides, but comprise only a small proportion of all epitopes. The minimal epitope size is about 5 amino acid residues. Also called a sequential epitope.

**LINKAGE** - The tendency of genes to be inherited together as a result of their relatively close proximity on the same chromosome, or location on the same plasmid.

**LINKER** - A short *oligodeoxyribonucleotide*, usually representing a specific *restriction endonuclease recognition sequence*, which may be *ligated* onto the termini of a DNA molecule to facilitate cloning. Following the ligation reaction, the product is digested with the endonuclease, generating a DNA fragment with the desired *sticky* or *blunt ends*.

**LIPOFECTIN** - A commercially marketed liposome suspension which is mixed with DNA or RNA to facilitate uptake of the nucleic acid by eukaryotic cells (see TRANSFECTION).

**LONG TERMINAL REPEATS (LTR)** - Nucleic acid sequences DNA that repeat hundreds or thousands of times. They are found in retroviral DNA and in retrotransposons flanking functional genes. They are used by viruses to insert their genetic sequences into the host genomes. The LTRs are partially transcribed into an RNA intermediate, followed by reverse transcription into complementary DNA (cDNA) and ultimately dsDNA (double-stranded DNA) with full LTRs.

The LTRs then mediate integration of the retroviral DNA copy via an LTR specific integrase into another region of the host chromosome.

**MELTING** - The dissociation of a duplex nucleic acid molecule into single strands, usually by increasing temperature. See DENATURATION.

**MISSENSE MUTATION** - A nucleotide mutation which results in a change in the amino acid sequence of the encoded protein (contrast with SILENT MUTATION).

**MOBILITY SHIFT** - See GEL SHIFT.

**MONOCLONAL ANTIBODY** - An antibody with very specific and often unique binding specificity which is secreted by a biologically cloned line of plasmacytoma cells in the absence of other related antibodies with different binding specificities. Differs from *polyclonal antibodies*, which are mixed populations of antibody molecules such as may be present in a serum specimen, within which many different individual antibodies have different binding specificities.

**MOTIF** - A recurring pattern of short sequence of DNA, RNA, or protein, that usually serves as a recognition site or active site. The same motif can be found in a variety of types of organisms.

**mRNA** - A cytoplasmic RNA which serves directly as the source of code for protein synthesis. See TRANSLATION.

**MULTICISTRONIC MESSAGE** - An *mRNA transcript* with more than one *cistron* and thus encoding more than one *polypeptide*. These generally do not occur in eukaryotic organisms, due to differences in the mechanism of translation initiation.

**MULTICOPY PLASMIDS** - Present in bacteria at amounts greater than one per chromosome. Vectors for cloning DNA are usually multicopy; there are sometimes advantages in using a single copy plasmid.

**MULTIPLE CLONING SITE** - An artificially constructed region within a *vector* molecule which contains a number of closely spaced *recognition sequences* for *restriction endonucleases*. This serves as a convenient site into which foreign DNA may be inserted.

**N TERMINUS** - See AMINO TERMINUS.

**NATIVE GEL** - An electrophoresis gel run under conditions which do not denature proteins (i.e., in the absence of SDS, urea, 2-mercaptoethanol, etc.).

**NESTED PCR** - A very sensitive method for amplification of DNA, which takes part of the product of a single *PCR* reaction (after 30-35 cycles), and subjects it to a new round of PCR using a different set of *PCR primers* which are nested within the region flanked by the original primer pair (see POLYMERASE CHAIN REACTION).

**NICK** - In *duplex* DNA, this refers to the absence of a *phosphodiester bond* between two adjacent *nucleotides* on one strand.

**NICK TRANSLATION** - A method for introducing labeled *nucleotides* into a double-stranded DNA molecule which involves making small *nicks* in one strand with DNase, and then repairing with DNA polymerase I.

**NONCONSERVATIVE SUBSTITUTION** - A mutation which results in the substitution of one *amino acid* within a polypeptide chain with an amino acid belonging to a different polarity/charge group (see AMINO ACIDS, CONSERVATIVE MUTATION)

**NONSENSE CODON** - See STOP CODON.

**NONSENSE MUTATION** - A change in the sequence of a nucleic acid that causes a *nonsense (stop or termination) codon* to replace a codon representing an amino acid.

**NONTRANSLATED RNA (NTR)** - The segments located at the 5' and 3' ends of a mRNA molecule which do not encode any part of the polyprotein; may contain important translational control elements.

**NORTHERN BLOT** - RNA molecules are separated by electrophoresis (usually in an agarose gel) on the basis of size, then transferred to a solid-phase support (nitrocellulose paper or suitable other membrane) and detected by *hybridization* with a labeled probe (see SOUTHERN BLOT, WESTERN BLOT).

**NUCLEOSIDE** - The composite sugar and *purine or pyrimidine base* which are present in *nucleotides* which are the basic building blocks of DNA and RNA. Compare with NUCLEOTIDE:

**Nucleoside** = Base + Sugar

**NUCLEOTIDE** - The composite phosphate, sugar, and *purine or pyrimidine base* which are the basic building blocks of the nucleic acids DNA and RNA. The five nucleotides are adenylic acid, guanylic acid (contain *purine bases*), and cytidylic acid, thymidylic acid, and uridylic acid (contain *pyrimidine bases*).

**Nucleotide** = Base + Sugar + Phosphate (1, 2, or 3)

**OLIGODEOXYRIBONUCLEOTIDE** - A short, single-stranded DNA molecule, generally 15-50 *nucleotides* in length, which may be used as a *primer* or a *hybridization probe*. Oligodeoxyribonucleotides are synthesized chemically under automated conditions.

**OLIGONUCLEOTIDE** - See OLIGODEOXYRIBONUCLEOTIDE.

**ONCOGENE** - One of a number of genes believed to be associated with the malignant transformation of cells; originally identified in certain oncogenic retroviruses (*v-onc*) but also present in cells (*c-onc*). See PROTO-ONCOGENE.

**OPEN READING FRAME** - A region within a reading frame of an mRNA molecule that potentially encodes a *polypeptide*; and which does not contain a *translational stop codon* (see READING FRAME).

**OPERATOR** - The site on DNA at which a *repressor* protein binds to prevent *transcription* from initiating at the adjacent *promoter*.

**OPERON** - A complete unit of bacterial gene expression and regulation, including the structural gene or genes, regulator gene(s), and control elements in DNA recognized by regulator gene products(s).

**ORIGIN** - A site within a DNA sequence of a chromosome, plasmid, or non-integrated virus at which replication of the DNA is initiated.

**OVERHANG** - A terminus of a *duplex* DNA molecule which has one or more unpaired nucleotides in one of the two strands (hence either a 3' or 5' overhang). Cleavage of DNA with many restriction endonucleases leaves such overhangs (see STICKY END).

**PACKAGE** - In recombinant DNA procedures, refers to the step of incorporation of *cosmid* or other lambda *vector* DNA with an *insert* into a *phage* head for transduction of DNA into host. Also refers to the incorporation of a specifically deleted or debilitated viral vector with a transgene into a coat protein envelope for introduction of the vector and transgene into a host cell. The gene for the particular coat protein is carried in a "packaging cell". The coated viral vector with transgene is secreted into the media surrounding the "packaging cell".

**PALINDROMIC SEQUENCE** - A nucleotide sequence which is the same when read in either direction, usually consisting of adjacent inverted repeats. *Restriction endonuclease recognition sites* are palindromes:

5'-->3'  
GAATTC *EcoRI* recognition site  
CTTAAG  
3'<--5'

**PCR** - See POLYMERASE CHAIN REACTION

**PEPTIDE** - A chain formed by two or more *amino acids* linked through *peptide bonds*: dipeptide = two *amino acids*, oligopeptide = small number of *amino acids*, etc.

**PEPTIDE** - A molecule formed by peptide bonds covalently linking two or more *amino acids*. Short peptides (generally less than 60 amino acid residues, and usually only half that length) can be chemically synthesized by one of several different methods; larger peptides (more correctly, *polypeptides*) are usually *expressed* from recombinant DNA.

**PEPTIDE BOND** - A covalent bond between two *amino acids*, in which the carboxyl group of one amino acid (X1--COOH) and the amino group of an adjacent amino acid (NH2--X2) react to form X1-CO-NH-X2 plus H<sub>2</sub>O.

**PHAGE** - See BACTERIOPHAGE.

**PHENOTYPE** - The appearance of other characteristics of an organism resulting from the interaction of its genetic constitution with the environment.

**PHOSPHATASE, ALKALINE** - An enzyme which catalyzes the hydrolysis of phosphomonoesters of the 5' nucleotides. Used to dephosphorylate (remove phosphate groups from) the 5' ends of DNA or RNA molecules, to facilitate 5' end-labeling with <sup>32</sup>P added back by T4 polynucleotide kinase; or to dephosphorylate the 5' ends of DNA molecules to prevent unwanted *ligation* reactions during cloning.

**PHOSPHODIESTER BOND** - The covalent bond between the 3' hydroxyl in the sugar ring of one nucleotide and the 5' phosphate group of the sugar ring of the adjacent nucleotide residue within a nucleic acid:

5'-Ribose- 3' - O - P(O)<sub>2</sub> - O - 5' -Ribose - 3' - etc.

**PHOSPHORYLATION** - The addition of a phosphate monoester to a macromolecule, catalyzed by a specific kinase enzyme. With respect to proteins, certain *amino acid* side chains (serine, threonine, tyrosine) are subject to phosphorylation catalyzed by protein kinases; altering the phosphorylation status of a protein may have dramatic effects on its biologic properties, and is a common cellular control mechanism. With respect to DNA, 5' ends must be phosphorylated for *ligation*.

**PLASMID** - An extrachromosomal, usually circular, double-stranded DNA which is capable of *replication* within a cell, and which usually contains and expresses genes encoding resistance to antibiotics. By strict definition, a plasmid is not essential to the life of the cell.

**PLAQUE FORMING UNITS** – pfu, a measure of the number of viable viruses in a suspension that are able to infect and lyse indicator cells to form a clear zone (the plaque).

**POINT MUTATION** - A single nucleotide substitution within a gene; there may be several point mutations within a single gene. Point mutations do not lead to a shift in reading frames, thus at most cause only a single amino acid substitution (see FRAMESHIFT MUTATION).

**POLY-A TRACK** - A lengthy adenylic acid polymer (RNA) which is covalently linked to the 3' end of newly synthesized *mRNA* molecules in the nucleus. Function not clear.

**POLYMERASE CHAIN REACTION (PCR)** - A DNA amplification reaction involving multiple (30 or more) cycles of *primer annealing*, *extension*, and denaturation, usually using a heat-stable DNA *polymerase* such as *Taq polymerase*. Paired primers are used, which are complementary to opposing strands of the DNA and which flank the area to be amplified. Under optimal conditions, single DNA sequence can be amplified a million-fold.

**POLYMORPHISM** - Variation within a DNA or RNA sequence.

**POLYNUCLEOTIDE KINASE** - Enzyme which catalyzes the transfer of the terminal phosphate of ATP to 5' hydroxyl termini of polynucleotides, either DNA or RNA. Usually derived from T4 bacteriophage.

**POLYPEPTIDE** - See PEPTIDE.

**pre-mRNA** - An RNA molecule which is *transcribed* from chromosomal DNA in the nucleus of eukaryotic cells, and subsequently processed through *splicing* reactions to generate the mRNA which directs protein synthesis in the cytoplasm.

**PRIMARY STRUCTURE** -Refers to the sequence of *amino acid residues* or *nucleotides* within protein or nucleic acid molecules, respectively (also see SECONDARY and TERTIARY STRUCTURE).

**PRIMER** - An *oligonucleotide* which is *complementary* to a specific region within a DNA or RNA molecule, and which is used to prime (initiate) synthesis of a new strand of complementary DNA at that specific site, in a reaction or series of reactions catalyzed by a *DNA polymerase*. The newly synthesized DNA strand will contain the primer at its 5' end. Typically, primers are chemically synthesized oligonucleotides 15-50 nucleotides in length, selected on the basis of a known sequence. However, "random primers" (shorter oligonucleotides, about 6 nucleotides in length, and comprising all possible sequences) may be used to prime DNA synthesis from DNA or RNA of unknown sequence. completely known, but probably serves to enhance stability of the RNA. Is frequently used to select mRNA for cloning purposes by *annealing* to a column containing a matrix bound to poly-uridylic acid.

**POLYACRYLAMIDE GEL (PAGE)** - Used to separate proteins and smaller DNA fragments and *oligonucleotides* by electrophoresis. When run under conditions which denature proteins (i.e., in the presence of 2-mercaptoethanol, SDS, and possibly urea), molecules are separated primarily on the basis of size.

**POLYCLONAL ANTIBODY** - See MONOCLONAL ANTIBODY.

**POLYMERASE** - An enzyme which catalyzes the addition of a *nucleotide* to a nucleic acid molecule. There are a wide variety of RNA and DNA polymerases which have a wide range of specific activities and which operate optimally under different conditions. In general, all polymerases require *templates* upon which to build a new strand of DNA or RNA; however, DNA polymerases also require a *primer* to initiate the new strand, while RNA polymerases start synthesis at a specific *promoter* sequence.

**POST-TRANSLATIONAL MODIFICATION** - Modifications made to a polypeptide molecule after its initial synthesis, this includes proteolytic cleavages, *phosphorylation*, *glycosylation*, carboxylation, addition of fatty acid moieties, etc.

**PRIMER EXTENSION** - A reaction in which DNA is *reverse transcribed* from an RNA *template* to which a specific oligonucleotide *primer* has been *annealed*. The new cDNA product is an extension of the primer, which is synthesized at the 3' end of the primer in a direction extending toward the 5' end of the RNA. This reaction is useful for exploring the extreme 5' end of RNA molecules.

**PROBE** - Usually refers to a DNA or RNA molecule which has been labeled with <sup>32</sup>P or with *biotin*, to facilitate its detection after it has specifically *hybridized* with a target DNA or RNA sequence. However, the term may also refer to antibody probes used in *western* blots.

**PROCESSING** - With respect to proteins, generally used to refer to proteolytic *post-translational modifications* of a polypeptide. In the case of RNA, processing may involve the addition of a 5' *cap* and 3' *poly-A* tracks as well as *splicing* reactions in the nucleus.

**PROCESSIVITY** - The extent to which an RNA or DNA *polymerase* adheres to a *template* before dissociating, determines the average length (in kilobases) of the newly synthesized nucleic acid strands. Also applies to the action of *exonucleases* in digesting from the ends to the middle of a nucleic acid.

**PROMOTER** - A specific sequence within a double-stranded DNA molecule that is recognized by an RNA *polymerase*, which binds to it and uses it to begin transcribing the DNA *template* into a new RNA. The location and orientation of the *promoter* within a DNA molecule determines the start site of the new RNA. Other proteins (e.g. transcriptional activators such as *sigma factor*) are usually required for an RNA polymerase to recognize a promoter (see TRANSCRIPTION).

**PROTO-ONCOGENE** - A cellular *oncogene*-like sequence which is thought to play a role in controlling normal cellular growth and differentiation.

**PSEUDOGENE** - Inactive but stable components of the genome which derived by duplication and mutation of an ancestral, active gene. Pseudogenes can serve as the donor sequence in *gene conversion* events.

**PSEUDOREVERTANT** - A mutant virus or organism which has recovered a wildtype phenotype due to a second-site mutation (potentially located in a different region of the genome, or involving a different polypeptide) which has eliminated the effect of the initial mutation.

**PSEUDOKNOT** - A feature of RNA *tertiary structure*; best visualized as two overlapping *stem-loops* in which the loop of the first stem-loop participates as half of the stem in the second stem-loop.

**PURINE BASES** - Adenine (A) or Guanine (G) (see NUCLEOTIDE).

**PULSED-FIELD GEL ELECTROPHORESIS (PFGE)** - Separation of large (>50 kb) pieces of DNA, including complete chromosomes and genomes, by rapidly alternating the direction of electrophoretic migration in agarose gels.

**PYRIMIDINE BASES** - Cytosine (C), Thymine (T) or Uracil (U) (see NUCLEOTIDE).

**READING FRAME** - Refers to a polypeptide sequence potentially encoded by a single-stranded *mRNA*. Because *codons* are nucleotide triplets, each mRNA has 3 reading frames (each nucleotide can participate in 3 codons, at the 1st, 2nd, and 3rd base position). Duplex DNA strands have 6 reading frames, 3 in each strand (see OPEN READING FRAME):

AlaSerProLeuVal . . 1st reading frame  
ProAlaProTERTrp . . 2nd reading frame: TER = Stop  
GlnProProSerGly . . 3rd reading frame

GCCAGCCCCCUAGTGGG... Nucleotide sequence of mRNA

**RECOGNITION SEQUENCE** - A specific *palindromic sequence* within a double-stranded DNA molecule which is recognized by a *restriction endonuclease*, and at which the restriction endonuclease specifically cleaves the DNA molecule.

**RECOMBINATION** - See HOMOLOGOUS RECOMBINATION.

**RECOMBINATION-REPAIR** - A mode of filling a gap in one strand of duplex DNA by retrieving a homologous single strand from another duplex. Usually the underlying mechanism behind *homologous recombination* and *gene conversion*.

**RELAXED DNA** - See SUPERCOIL.

**REPLICATION** - The copying of a nucleic acid molecule into a new nucleic acid molecule of similar type (i.e., DNA --> DNA, or RNA --> RNA).

**REPORTER GENE** - The use of a functional enzyme, such as beta-galactosidase, luciferase, or chloramphenicol acetyltransferase, downstream of a gene, promoter, or translational control element of interest, to more easily identify successful introduction of the gene into a host and to measure transcription and/or translation.

**REPRESSION** - Inhibition of transcription (or translation) by the binding of a repressor protein to a specific site on DNA (or *mRNA*).

**RESIDUE** - As applied to proteins, what remains of an *amino acid* after its incorporation into a peptide chain, with subsequent loss of a water molecule (see PEPTIDE BOND).

**RESTRICTION ENDONUCLEASE** - A bacterial enzyme which recognizes a specific *palindromic sequence (recognition sequence)* within a double-stranded DNA molecule and then catalyzes the cleavage of both strands at that site. Also called a restriction enzyme. Restriction endonucleases may generate either *blunt* or *sticky ends* at the site of cleavage.

**RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP)** - Variations in the lengths of fragments of DNA generated by digestion of different DNAs with a specific *restriction endonuclease*, reflecting genetic variation (*polymorphism*) in the DNAs.

**RESTRICTION FRAGMENTS** - DNA fragments generated by digestion of a DNA preparation with one or more *restriction endonucleases*; usually separated by *agarose gel electrophoresis* and visualized by *ethidium bromide* staining under UV light (or alternatively subjected to *Southern blot analysis*).

**RESTRICTION MAP** - A linear array of sites on a particular DNA which are cleaved by various selected *restriction endonucleases*.

**RESTRICTION SITE** - See RECOGNITION SEQUENCE.

**RETICULOCYTE LYSATE** - A lysate of rabbit reticulocytes, which has been extensively digested with micrococcal nuclease to destroy the reticulocyte *mRNAs*. With the addition of an exogenous, usually synthetic, mRNA, *amino acids* and a source of energy (ATP), the translational machinery of the reticulocyte (*ribosomes*, eukaryotic translation factors, etc.) will permit *in vitro translation* of the

added mRNA with production of a new *polypeptide*. This is only one of several available in vitro translation systems.

**RETROVIRUS** - An RNA virus that is replicated in a host cell via the enzyme reverse transcriptase to produce DNA from its RNA genome. The DNA is then incorporated into the host's genome by an integrase enzyme. The virus thereafter replicates as part of the host cell's DNA. Retroviruses are enveloped viruses that belong to the viral family *Retroviridae*. A special variant of retroviruses are endogenous retroviruses which are integrated into the genome of the host and inherited across generations.

The virus itself stores its nucleic acid in the form of a +mRNA (including the 5'cap and 3'PolyA inside the virion) genome and serves as a means of delivery of that genome into cells it targets as an obligate parasite, and constitutes the infection. Once in the host's cell, the RNA strands undergo reverse transcription in the cytoplasm and are integrated into the host's genome, at which point the retroviral DNA is referred to as a provirus. It is difficult to detect the virus until it has infected the host. In most viruses, DNA is transcribed into RNA, and then RNA is translated into protein. However, retroviruses function differently - their RNA is reverse-transcribed into DNA, which is integrated into the host cell's genome (when it becomes a provirus), and then undergoes the usual transcription and translational processes to express the genes carried by the virus. Therefore, the order of steps from a retroviral gene to a retroviral protein is: RNA → DNA → RNA → Protein.

**REVERSE TRANSCRIPTASE** - A DNA polymerase which copies an RNA molecule into single-stranded cDNA; usually purified from retroviruses.

**REVERSE TRANSCRIPTION** - Copying of an RNA molecule into a DNA molecule.

**REVERTANT** - See BACK MUTATION.

**RIBONUCLEASE (RNase)** - An enzyme which catalyzes the hydrolysis of RNA. There are many different RNases, some of the more important include:

RNase A Cleaves ssRNA 3' of pyrimidines  
RNase T1 Cleaves ssRNA at guanosine nucleotides  
RNase V1 Cleaves dsRNA (helical regions)  
RNase H Degrades the RNA part of RNA:DNA hybrids.

**RIBOSOMAL BINDING SEQUENCE** (Shine-Dalgarno sequence) - In prokaryotic organisms, part or all of the polypurine sequence AGGAGG located on *mRNA* just upstream of an AUG *initiation codon*; it is complementary to the sequence at the 3' end of 16S rRNA; and involved in binding of the ribosome to *mRNA*. The *internal ribosomal entry site* found in some viruses may be an analogous eukaryotic genetic element.

**RIBOSOME** - A complex ribonucleoprotein particle (eukaryotic ribosomes contain 4 RNAs and at least 82 proteins) which is the "machine" which translates *mRNA* into protein molecules. In eukaryotic cells, ribosomes are often in close proximity to the endoplasmic reticulum.

**RIBOZYME** - A catalytically active RNA. A good example is the hepatitis delta virus RNA which is capable of self-cleavage and self-ligation in the absence of protein enzymes.

**RNA POLYMERASE** - A polymerase which synthesizes RNA (see POLYMERASE).

**RNA SPLICING** - A complex and incompletely understood series of reactions occurring in the nucleus of eukaryotic cells in which *pre-mRNA transcribed* from chromosomal DNA is processed such that noncoding regions of the pre-mRNA (*introns*) are excised, and coding regions (*exons*) are covalently linked to produce an *mRNA* molecule ready for transport to the cytoplasm. Because of splicing,



eukaryotic DNA representing a gene encoding any given protein is usually much larger than the mRNA from which the protein is actually *translated*.

**RNase** - see RIBONUCLEASE

**rRNA** - Ribosomal RNA (four sizes in humans: 5S, 5.8S, 18S, and 28S); RNA component of the *ribosome*, which may play catalytic roles in *translation*.

**RT/PCR REACTION** - A series of reactions which result in RNA being copied into DNA and then amplified. A single *primer* is used to make single-stranded *cDNA* copies from an RNA *template* under direction of *reverse transcriptase*. A second primer *complementary* to this "first strand" cDNA is added to the reaction mix along with *Taq polymerase*, resulting in synthesis of double-stranded DNA. The reaction mix is then cycled (denaturation, *annealing* of primers, *extension*) to amplify the DNA by conventional PCR.

**RUNOFF TRANSCRIPT** - RNA which has been synthesized from plasmid DNA (usually by a bacteriophage *RNA polymerase* such as *T7* or *SP6*) and which terminates at a specific 3' site because of prior cleavage of the plasmid DNA with a *restriction endonuclease*.

**S1 NUCLEASE** - An enzyme which digests single-stranded DNA or RNA

**SDS-PAGE** - Sodium Dodecyl Sulphate detergent is added into gels and buffers to denature the protein. Denaturing protein gel electrophoresis (see POLYACRYLAMIDE GEL ELECTROPHORESIS).

**SECONDARY STRUCTURE** - (also see PRIMARY and TERTIARY STRUCTURE) Local structure within a protein which is conferred by the nature of the side chains of adjacent *amino acids* (e.g., alpha helix, beta sheet, random coil); local structure within an RNA molecule which is conferred by *base pairing* of *nucleotides* which are relatively closely positioned within the sequence (e.g., hairpins, stem-loop structures).

**SELECTION** - The use of particular conditions, such as the presence of ampicillin, to allow survival only of cells with a particular *phenotype*, such as production of beta-lactamase.

**SEQUENCE POLYMORPHISM** - See POLYMORPHISM.

**SEQUENTIAL EPITOPE** - See LINEAR EPITOPE.

**SHOTGUN CLONING or SEQUENCING** - Cloning of an entire genome or large piece of DNA in the form of randomly generated small fragments. The individual sequences obtained from the clones will be used to construct *contigs*.

**SHUTTLE VECTOR** - A small *plasmid* capable of *transfection* into both prokaryotic and eukaryotic cells.

**SIDE CHAIN** - See AMINO ACID.

**SIGMA FACTOR** - Certain small ancillary proteins in bacteria that increase the binding affinity of RNA polymerase to a promoter. Different sigma factors recognize different promoter sequences.

**SIGNAL SEQUENCE** - A hydrophobic amino acid sequence which directs a growing peptide chain to be secreted into the endoplasmic reticulum.

**SILENT MUTATION** - A nucleotide substitution (never a single deletion or insertion) which does not alter the amino acid sequence of an encoded protein due to the *degeneracy* of the genetic code. Such mutations usually involve the third base (*wobble* position) of *codons*.

**SITE-DIRECTED MUTAGENESIS** - The introduction of a mutation, usually a *point mutation* or an insertion, into a particular location in a cloned DNA fragment. This mutated fragment may be used to "knock out" a gene in the organism of interest by *homologous recombination*.

**SITE-SPECIFIC RECOMBINATION** - Occurs between two specific but not necessarily homologous sequences. Usually catalyzed by enzymes not involved in general or *homologous recombination*.

**SOUTHERN BLOT** - DNA is separated by electrophoresis (usually in *agarose gels*), then transferred to nitrocellulose paper or other suitable solid-phase matrix (e.g., nylon membrane), and denatured into single strands so that it can be *hybridized* with a specific *probe*. The Southern blot was developed by E.M. Southern, a molecular biologist in Edinburgh. *Northern* and *western* blots were given contrasting names to reflect the different target substances (RNA and proteins, respectively) that are subjected in these procedures to electrophoresis, blotting and subsequent detection with specific probes.

**SOUTHWESTERN BLOT** - The binding of protein to a nucleic acid on a matrix similar to what is done for western, northern, and southern blots. This technique is used to identify DNA binding proteins and the recognition sites for these proteins.

**SP6 RNA POLYMERASE** - A bacteriophage *RNA polymerase* which is commonly used to transcribe *plasmid* DNA into RNA. The plasmid must contain an SP6 *promoter* upstream of the relevant sequence.

**SPLICING** - see RNA SPLICING.

**ss** - Single stranded.

**START CODON** - See INITIATION CODON.

**STEM-LOOP** - A feature of RNA *secondary structure*, in which two complementary, inverted sequences which are separated by a short-intervening sequence within a single strand of RNA base pair to form a "stem" with a "loop" at one end. Similar to a *hairpin*, but these usually have very small loops and longer stems.

**STICKY END** - The terminus of a DNA molecule which has either a 3' or 5' overhang, and which typically results from a cut by a *restriction endonuclease*. Such termini are capable of specific ligation reactions with other termini which have complementary overhangs. A sticky end can be "blunt ended" either by the removal of an overhang, or a "filling in" reaction which adds additional nucleotides complementary to the overhang (see BLUNT END).

5'-->3'  
NNNG AATTCNNN *EcoRI* cut, 5' overhang  
NNNCTTAA GNNN  
3'<--5'

5'-->3'  
XXXAGCGC TNNN *HaeII* cut, 3' overhang  
XXXT CGCGANNN  
3'<--5'

**STOP CODON** - A *codon* (UAA, UAG, UGA) which terminates *translation*.

**STREPTAVIDIN** - A bacterial analog of egg white *avidin*.

**STRINGENCY** - The conditions employed for *hybridization* which determine the specificity of the *annealing* reaction between two single-stranded nucleic acid molecules. Increasingly stringent conditions may be reached by raising temperature or lowering ionic strength, resulting in greater specificity (but lower sensitivity) of the hybridization reaction.

**SUPERCOIL** - Double-stranded circular DNA which is twisted about itself. Commonly observed with *plasmids* and circular viral DNA genomes (such as that of hepatitis B virus). A nick in one strand of the plasmid may remove the twist, resulting in a *relaxed*, circular DNA molecule. A complete break in the DNA puts the plasmid in a linear form. Supercoils, relaxed circular DNA, and linear DNA all have different migration properties in agarose gels, even though they contain the same number of base pairs.

**T7 RNA POLYMERASE** - A bacteriophage *RNA polymerase* which is commonly used to transcribe plasmid DNA into RNA. The plasmid must contain a T7 *promoter* upstream of the relevant sequence.

**Taq POLYMERASE** - A *DNA polymerase* which is very stable at high temperatures, isolated from the thermophilic bacterium *Thermus aquaticus*. Very useful in *PCR* reactions which must cycle repetitively through high temperatures during the denaturation step.

**TEMPLATE** - A nucleic acid strand, upon which a *primer* has *annealed* and a nascent RNA strand is being extended.

**TERMINATION CODON** - See STOP CODON.

**TERMINATOR** - A sequence *downstream* from the 3' end of an *open reading frame* that serves to halt *transcription* by the RNA polymerase. In bacteria these are commonly sequences that are *palindromic* and thus capable of forming *hairpins*. Sometimes termination requires the action of a protein, such as Rho factor in *E. coli*.

**TERTIARY STRUCTURE** - (also see PRIMARY and SECONDARY STRUCTURE) Refers to higher ordered structures conferred on proteins or nucleic acids by interactions between *amino acid residues* or *nucleotides* which are not closely positioned within the sequence (primary structure) of the molecule.

**Tm** - The midpoint of the temperature range over which DNA is melted or denatured by heat; the temperature at which a *duplex* nucleic acid molecule is 50% *melted* into single strands, it is dependent upon the number and proportion of G-C *base pairs* as well as the ionic conditions. Often referred to as a measure of the thermal stability of a nucleic acid *probe:target* sequence hybrid.

**TRANS** - As used in molecular biology, an interaction that involves two sites which are located on separate molecules.

**TRANSCRIPT** - A newly made RNA molecule which has been copied from DNA.

**Transcription** - The copying of a DNA template into a single-stranded RNA molecule. The processes whereby the transcriptional activity of eukaryotic genes are regulated are complex, involve a variety of accessory transcriptional factors which interact with *promoters* and *polymerases*, and constitute one of the most important areas of biological research today.

**TRANSCRIPTION/TRANSLATION REACTION** - An *in vitro* series of reactions, involving the synthesis (*transcription*) of an mRNA from a *plasmid* (usually with T7 or SP6 *RNA polymerase*), followed by use of the mRNA to program *translation* in a cell-free system such as a rabbit *reticulocyte lysate*. The *polypeptide* product of translation is usually labelled with [35S]-methionine, and examined

in an *SDS-PAGE* gel with or without prior *immunoprecipitation*. This series of reactions permits the synthesis of a polypeptide from DNA *in vitro*.

**TRANSCRIPTIONAL START SITE** - The nucleotide of a gene or cistron at which *transcription* (RNA synthesis) starts; the most common triplet at which transcription begins in *E. coli* is CAT. *Primer extension* identifies the transcriptional start site.

**TRANSFECTION** - The process of introducing foreign DNA (or RNA) into a host organism, usually a eukaryotic cell.

**TRANSFORMATION** - Multiple meanings. With respect to cloning of DNA, refers to the transformation of bacteria (usually to specific antibiotic resistance) due to the uptake of foreign DNA. With respect to eukaryotic cells, usually means conversion to less-restrained or unrestrained growth.

**TRANSGENE** - A foreign gene which has been introduced into the germ line of an animal species.

**TRANSGENIC** - An animal (usually a mouse) or plant into which a foreign gene has been introduced in the germ line. An example: transgenic mice expressing the human receptor for poliovirus are susceptible to human polioviruses.

**TRANSITION** - A nucleotide substitution in which one pyrimidine is replaced by the other pyrimidine, or one purine replaced by the other purine (e.g., A is changed to G, or C is changed to T) (contrast with TRANSVERSION) .

**TRANSLATION** - The process whereby *mRNA* directs the synthesis of a protein molecule; carried out by the *ribosome* in association with a host of translation initiation, elongation and termination factors. Eukaryotic genes may be regulated at the level of translation, as well as the level of *transcription*.

**TRANSLOCATION** - The process by which a newly synthesized protein is directed toward a specific cellular compartment (i.e, the nucleus, the endoplasmic reticulum).

**TRANSPOSON** - A transposable genetic element; certain sequence elements which are capable of moving from one site to another in a DNA molecule without any requirement for sequence relatedness at the donor and acceptor sites. Many transposons carry antibiotic resistance determinants and have *insertion sequences* at both ends, and thus have two sets of *inverted repeats*.

**TRANSPOSITION** - The movement of DNA from one location to another location on the same molecule, or a different molecule within a cell.

**TRANSVERSION** - A nucleotide substitution in which a purine replaces a pyrimidine, or vice versa (e.g., A is changed to T, or T is changed to G) (see TRANSITION)

**TRIPLET** - A three-nucleotide sequence; a *codon*.

**tRNA** - Small, tightly folded RNA molecules which act to bring specific amino acids into *translationally* active *ribosomes* in a fashion which is dependent upon the *mRNA* sequence. One end of the tRNA molecule recognizes the nucleotide triplet which is the *codon* of the mRNA, while the other end (when activated) is covalently linked to the relevant *amino acid*.

**UNTRANSLATED REGION** - portion of RNA or DNA that is not translated into protein. The region, usually at the end of an mRNA does not code for an amino acid but may be involved in regulation of translation.

**UNTRANSLATED RNA** - See NONTRANSLATED RNA.

**UPSTREAM** - Identifies sequences located in a direction opposite to that of expression; for example, the bacterial *promoter* is upstream of the *initiation codon*. In an mRNA molecule, upstream means toward the 5' end of the molecule. Occasionally used to refer to a region of a polypeptide chain which is located toward the amino terminus of the molecule.

**VECTOR** - A *plasmid*, *cosmid*, bacteriophage, or virus which carried foreign nucleic acid into a host organism.

**WESTERN BLOT** - Proteins are separated by *SDS-PAGE*, then electrophoretically transferred to a solid-phase matrix such as nitrocellulose, then probed with a labelled antibody (or a series of antibodies)

**WILDTYPE** - The native or predominant genetic constitution before mutations, usually referring to the genetic constitution normally existing in nature.

**WOBBLE POSITION** - The third base position within a *codon*, which can often (but not always) be altered to another nucleotide without changing the encoded amino acid (see DEGENERACY).