Glossary of Biochemistry and Molecular Biology

3' end/5' end: A nucleic acid strand is inherently directional, and the "5 prime end" has a free hydroxyl (or phosphate) on a 5' carbon and the "3 prime end" has a free hydroxyl (or phosphate) on a 3' carbon (carbon atoms in the sugar ring are numbered from 1' to 5'. That's simple enough for an RNA strand or for single-stranded (ss) DNA. However, for double-stranded (ds) DNA it's not so obvious - each strand has a 5' end and a 3' end, and the 5' end of one strand is paired with the 3' end of the other strand (it is "antiparallel"). One would talk about the 5' end of ds DNA only if there was some reason to emphasize one strand over the other - for example if one strand is the sense strand of a gene. In that case, the orientation of the sense strand establishes the direction.

3' flanking region: A region of DNA which is not copied into the mature mRNA, but which is present adjacent to 3' end of the gene. It was originally thought that the 3' flanking DNA was not transcribed at all, but it was discovered to be transcribed into RNA, but quickly removed during processing of the primary transcript to form the mature mRNA. The 3' flanking region often contains sequences which affect the formation of the 3' end of the message. It may also contain enhancers or other sites to which proteins may bind.

3' untranslated region: A region of the DNA which is transcribed into mRNA and becomes the 3' end or the message, but which does not contain protein coding sequence. Everything between the stop codon and the polyA tail is considered to be 3' untranslated (see Figure 4). The 3' untranslated region may affect the translation efficiency of the mRNA or the stability of the mRNA. It also has sequences which are required for the addition of the poly(A) tail to the message (including one known as the "hexanucleotide", AAUAAA).

5' flanking region: A region of DNA which is not transcribed into RNA, but rather is adjacent to 5' end of the gene. The 5'-flanking region contains the promoter, and may also contain enhancers or other protein binding sites.

5' untranslated region: A region of a gene which IS transcribed into mRNA, becoming the 5' end of the message, but which does not contain protein coding sequence. The 5'-untranslated region is the portion of the DNA starting from the cap site and extending to the base just before the ATG translation initiation codon. While not itself translated, this region may have sequences which alter the translation efficiency of the mRNA, or which affect the stability of the mRNA.

7-methyl-guanosine triphosphate: The methylated form of guanosine triphosphate used to cap an mRNA molecule in most eukaryotes. (See cap)

A

Ablation experiment: An experiment designed to produce an animal deficient in one or a few cell types, in order to study cell lineage or cell function. The idea is to make a transgenic mouse with a toxin gene (often diphtheria toxin) under control of a specialized promoter which activates only in the target cell type. When embryo development progresses to the point where it starts to form the target tissue, the toxin gene is activated, and that specific tissue dies. Other tissues are unaffected.

Acrylamide gels: A polymer gel used for electrophoresis of DNA or protein to measure their sizes (in daltons for proteins, or in base pairs for DNA). See "Gel electrophoresis". Acrylamide gels are especially useful for high-resolution separations of DNA in the range of tens to hundreds of nucleotides in length.

Adapter: A short chemically synthesized DNA double strand which can be used to link the ends of two DNA molecules.

Adenine (A): One of the purine bases found in DNA and RNA (6-aminopurine), one member of the base pair A- T (adenine- thymine).

Agarose gel electrophoresis: A method to analyze the size of DNA (or RNA) fragments. In the presence of an electric field, larger fragments of DNA move through a gel slower than smaller ones, producing different migrating "bands". Usually, these are visualized by soaking the gel in a dye (ethidium bromide) which makes the DNA fluoresce under UV light. This is the gel of choice for DNA or RNA in the range of thousands of bases in length, or even up to 1 megabase if you are using pulsed field gel electrophoresis. (See also electrophoresis).

Alkaline Phosphatase: 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 32P added back by T4 polynucleotide kinase; or to dephosphorylate the 5' ends of DNA molecules to prevent unwanted ligation reactions during cloning.

Allele: An allele is one of the alternative (two or more) forms of a particular gene inherited separately from each parent; usually found at the same locus on homologous chromosomes. Equivalent genes in the two sets might be different, for example because of single nucleotide polymorphisms.

Allele-specific ligation: Technique permitting discrimination of two allele at locus by providing two short synthetic oligonucleotides that would bind ajjacent to teach other on amplified DNA fragment, and would be ligated in presence of DNA ligase; if one of alleles containing mutation overlapped by 3'-end of one oligonucleotide, its ligation to oligonucleotide bound 3' to it would be prevented; to deduce identity of unknown allele, differentially labelled oligonucleotide pairs may be designed for each allele and their ligation efficiency compared in presence of unknown allele.

Allele-specific PCR (AS-PCR): Refers to amplification of specific alleles, or DNA sequence varients at same locus; specificity is achieved by designing one or both PCR primers so that they partially overlap site of sequence difference between amplified alleles.

Alternative splicing: Refers to fact that certain genes retain or omit particular exons in the final spliced transcript.

Alu sequence: A family of sequence-related elements about 300 bp in length approximately 500 000 copies of which are scattered along the human genome.

Amino acid: Any of a class of 20 molecules that are combined to form proteins in living things. Consisting of the basic formula NH2-CHR-COOH, where "R" is the side chain which defines the amino acid:The sequence of amino acids in a protein and hence protein function are determined by the genetic code.

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 NH2 end of a peptide chain (by custom drawn at the left of a protein sequence)

Amp resistance: See "Antibiotic resistance".

Amplification: An increase in the number of copies of a specific DNA fragment; can either be *in vivo* or *in vitro*. See cloning, polymerase chain reaction.

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.

Aneuploid: A cell containing a number of chromosomes that is not an even multiple of the haploid number (n).

Anneal: The act of two nucleic acid sequences hydrogen bonding through complementarity of the bases determining their sequences. Generally synonymous with "hybridize".

Antibiotic resistance: Resistance conferred to the host the ability to survive a given antibiotic by plasmids containing resistance genes. If such a plasmid is present in a host, that host will not be killed by (moderate levels of) ampicillin or tetracycline. By transforming plasmids containing antibiotic resistance genes, one can get rid of all the contaminating bacteria which does not have such plasmid, thus ensuring that the plasmid will be propagated as the surviving cells divide.

Antibody: An immunoglobulin protein produced by B-lymphocytes of the immune system that binds to a specific antigen molecule. (See monoclonal antibodies, polyclonal antibodies.)

Anticodon: The 3-base sequence of a tRNA that base-pairs with the mRNA codon to effect translation of the mRNA into polypeptide sequence. (See transfer RNA and messanger RNA).

Antigen: Any foreign substance, such as a virus, bacterium, or protein that elicits an immune response by stimulating the production of antibodies.

Antigenic variation: Mechanism to ensure rapid sequence variation of the gene(s) encoding homologues of an individual protein antigen; usually involving multiple, related gene copies.

Antisense construct: Plasmid that contains coding region of gene placed adjacent to promoter in such orientation as to cause gene to be transcribed from DNA strand complementary to that from which RNA is normally transcribed.

Anti-sense strand: See Sense strand.

AP-1 site: The binding site on DNA at which the transcription "factor" AP-1 binds, thereby altering the rate of transcription for the adjacent gene. AP-1 is actually a complex between c-fos protein and c-jun protein, or sometimes is just c-jun dimers. The AP-1 site consensus sequence is (C/G) TGACT(C/A) A. Also known as the TPA-response element (TRE). (TPA is a phorbol ester, tetradecanoyl phorbol acetate, which is a chemical tumor promoter).

Arrayed library: Individual primary recombinant clones (hosted in phage, cosmid, YAC, or other vector) that are placed in two- dimensional arrays in microtiter dishes. Each primary clone can be identified by the identity of the plate and the clone location (row and column) on that plate. Arrayed libraries of clones can be used for many applications, including screening for a specific gene or genomic region of interest as well as for physical mapping. Information gathered on individual clones from various genetic linkage and physical map analyses is entered into a relational database and used to construct physical and genetic linkage maps simultaneously; clone identifiers serve to interrelate the multilevel maps. Compare library, genomic library.

ATG or AUG: The codon for methionine; the translation initiation codon. Usually, protein translation can only start at a methionine codon (although this codon may be found elsewhere within the protein sequence as well). In eukaryotic DNA, the sequence is ATG; in RNA it is AUG. Usually, the first AUG in the mRNA is the point at which translation starts, and an open reading frame follows - i.e. the nucleotides taken three at a time will code for the amino acids of the protein, and a stop codon will be found only when the protein coding region is complete.

Attenuation: A form of gene regulation wherein termination of transcription is controlled to regulate overall levels of gene expression.

Avidin: A glycoprotein which binds to biotin with very high affinity (Kd = 10-15).

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.

Autosome: A chromosome not involved in sex determination. The diploid human genome consists of 46 chromosomes, 22 pairs of autosomes, and 1 pair of sex chromosomes (the X and Y chromosomes).

B

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).

Bacterial artificial chromosome (BAC): A chromosome-like structure, constructed by genetic engineering. BAC is a cloning vector capable of carrying between 100 and 300 kilobases of target sequence. They are propagated as a mini-chromosome in a bacterial host. The size of the typical BAC is ideal for use as an intermediate in large-scale genome sequencing projects. Entire genomes can be cloned into BAC libraries, and entire BAC clones can be shotgun-sequenced fairly rapidly.

Bacteriophage: (simply phage) A virus that infects a bacterium and which is often used in molecular genetics experiments as a vector, or cloning vehicle. Recombinant phages can be made in which certain non-essential l DNA is removed and replaced with the DNA of interest. The phage can accommodate a DNA "insert" of about 15-20 kb. Replication of that virus will thus replicate the investigator's DNA. One would use phage l rather than a plasmid if the desired piece of DNA is rather large. Bacteriophage l and M13 are ones commonly used in cloning and/or subcloning of small genes or DNA fragments in *E. coli.* Bacteriophage P1 is one that is used for fragments up to 95 Kb in size.

Band: A pattern of light and dark regions by Giemsa staining that can serve as landmarks on chromosomes.

Band shift assay: see Gel shift assay.

Base: In molecular biology, this term refers to the purine bases adenine and guanine, and the pyrimidine bases uracil, thymine, and cytosine, or modification of these bases.

Base Pair (bp): 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.

Base sequence: The order of nucleotide bases in a DNA molecule.

Base sequence analysis: A method, sometimes automated, for determining the base sequence.

B-DNA: Conformation of the Watson-Crick double helix in which the two strands form a right-handed double helix.

Binding site: A place on cellular DNA to which a protein (such as a transcription factor) can bind. Typically, binding sites might be found in the vicinity of genes, and would be involved in activating transcription of that gene (promoter elements), in enhancing the transcription of that gene (enhancer elements), or in reducing the transcription of that gene (silencers). NOTE that whether the protein in fact performs these functions may depend on some condition, such as the presence of a hormone, or the tissue in which the gene is being examined. Binding sites could also be involved in the regulation of chromosome structure or of DNA replication.

Biotechnology: A set of biological techniques developed through basic research and now applied to research and product development. In particular, the use by industry of recombinant DNA, cell fusion, and new bioprocessing techniques.

Biotin: A coenzyme which is essential for carboxylation reactions (see Avidin).

Blotting: A technique for detecting one RNA within a mixture of RNAs (a Northern blot) or one type of DNA within of a mixture of DNAs (a Southern blot). A blot can prove whether that one species of RNA or DNA is present, how much is there, and its approximate size. Basically, blotting involves gel electrophoresis, transfer to a blotting membrane (typically nitrocellulose or activated nylon), and incubating with a radioactive probe. Exposing the membrane to X-ray film produces darkening at a spot correlating with the position of the DNA or RNA of interest. The darker the spot, the more nucleic acid was present there.

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 SmaI cut, no overhang

NNNGGG CCCNNN

3'<--5'

bp: See 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

C Terminus: See Carboxyl Terminus.

Cancer: A disease process initiated by transformation of a cell to behavior lacking in normal proliferative control, and often involving invasive behavior and localized vascularization.

Cap: All eukaryotes have at the 5' end of their messages a structure called a "cap", consisting of a 7-methylguanosine in 5'-5' triphosphate linkage with the first nucleotide of the mRNA. It is added post-transcriptionally, and is not encoded in the DNA.

Cap site: Two usages:

1] In eukaryotes, the cap site is the position in the gene at which transcription starts, and really should be called the "transcription initiation site". The first nucleotide is transcribed from this site to start the nascent RNA chain. That nucleotide becomes the 5' end of the chain, and thus the nucleotide to which the cap structure is attached (see "Cap").

2] In bacteria, the CAP site (note the capital letters) is a site on the DNA to which a protein factor (the Catabolite Activated Protein) binds.

Carboxyl Terminus: Refers to the COOH end of a peptide chain (by custom drawn at the right of a protein sequence)

CAT assay: An enzyme assay. CAT stands for chloramphenicol acetyl transferase, a bacterial enzyme which inactivates chloramphenicol by acetylating it. CAT assays are often performed to test the function of a promoter. The gene coding for CAT is linked onto a promoter (transcription control region) from another gene, and the construct is "transfected" into cultured cells. The amount of CAT enzyme produced is taken to indicate the transcriptional activity of the promoter (relative to other promoters which must be tested in parallel). It is easier to perform a CAT assay than it is to do a Northern blot, so CAT assays were a common method for testing the effects of sequence changes on promoter function. Largely supplanted by the reporter gene luciferase.

CCAAT box: (CAT box, CAAT box, and other variants) A sequence found in the 5' flanking region of certain genes which is necessary for efficient expression. A transcription factor (CCAAT-binding protein, CBP) binds to this site.

cDNA: See complementary DNA.

cDNA clone: "complementary DNA"; a piece of DNA copied from an mRNA. The term "clone" indicates that this cDNA has been spliced into a plasmid or other vector in order to propagate it. A cDNA clone may contain DNA copies of such typical mRNA regions as coding sequence, 5'-untranslated region, 3' untranslated region or poly(A) tail. No introns will be present, nor any promoter sequences (or other 5' or 3' flanking regions). A "full-length" cDNA clone is one which contains all of the mRNA sequence from nucleotide #1 through to the poly(A) tail.

Centimorgan (cM): A unit of measure of the statistical probability recombination frequency between alleles. One centimorgan is equal to a 1% chance that a marker at one genetic locus will be separated from a marker at a second locus due to crossing over in a single generation. In human beings, 1 centimorgan is equivalent, on average, to 1 million base pairs.

Central dogma: A phrase that refers to the concept of information flow proceeding only from DNA to RNA to protein.

Centromere: A specialized constricted region of a chromosome to which spindle fibers attach during cell division at which two sister chromatids (the two exact copies of each chromosome that are formed after replication) are joined, and which attach to the spindle during cell division.

Chaperone Proteins: A series of proteins present in the endoplasmic reticulum which prevent proteins from folding prematurely and guide the proper folding of secreted proteins through a complex series of binding and release reactions.

Chromatid: A single, continuous double stranded DNA molecule with its unique, complete grouping of genetic information, associated proteins, higher-order structures, and centromeric and telomeric regions necessary for separation and maintenance after replication.

Chromosomes: A condensed, fibrillar, self- replicating genetic structures of cells containing the cellular DNA that bears in its nucleotide sequence the linear array of genes. In prokaryotes, chromosomal DNA is circular, and the entire genome is carried on one chromosome. Eukaryotic genomes consist of a number of chromosomes whose DNA is associated with different kinds of proteins.

Chromosome jumping: A technique whereby one starts with a piece of DNA from one region of a chromosome, and obtains clones from nearby regions without cloning everything in between (as in chromosome walking; see below). One round of jumping yields new clones at distances of several tens of kb away from the starting point. In practice, this method is used when classical genetics proves that a known piece of DNA is located on the chromosome close to a gene you would like to clone (such as a human disease gene). By cloning fragments some distance away in both directions from the known fragment, one might obtain (1) fragments further from the desired gene (which are discarded); (2) fragments even more closely linked to the desired gene (in which case one goes for another round of jumping); or (3) fragments from within the desired gene - the optimal result.

Chromosome walking: A technique for cloning everything in the genome around a known piece of DNA (the starting probe). You screen a genomic library for all clones hybridizing with the probe, and then figure out which one extends furthest into the surrounding DNA. The most distal piece of this most distal clone is then used as a probe, so that ever more distal regions can be cloned. This has been used to move as much as 200 kb away from a given starting point (an immense undertaking). Typically used to "walk" from a starting point towards some nearby gene in order to clone that gene. Also used to obtain the remainder of a gene when you have isolated a part of it.

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.

Clones: A group of cells derived from a single ancestor.

Clone (verb):

Clone (verb):. The act of duplicating genetic material within a vector. To clone a piece of DNA, one would insert it into some type of vector (say, a plasmid) and put the resultant construct into a host (usually a bacterium) so that the plasmid and insert replicate with the host. An individual bacterium is isolated and grown and the plasmid containing the "cloned" DNA is re-isolated from the bacteria, at which point there will be many millions of copies of the DNA - essentially an unlimited supply. Actually, an investigator wishing to clone some gene or cDNA rarely has that DNA

in a purified form, so practically speaking, to "clone" something involves screening a cDNA or genomic library for the desired clone. See also "Probe" for a description of how one might start a cloning project, and "Screening" for how the probe in used.

One can also clone more complex organisms, with considerable difficulty. The much-publicized Scottish research that resulted in the sheep 'Dolly' exemplifies this approach.

Clone (noun): The term "clone" can refer either to a bacterium carrying a cloned DNA, or to the cloned DNA itself.

Clone bank: See genomic library.

Cloning: to allow it to be sequenced or studied in some other way.

Cloning: The process of generating sufficient copies of a particular piece of DNA or asexually producing a group of cells (clones), all genetically identical, from a single ancestor. In recombinant DNA technology, the use of DNA manipulation procedures to produce multiple copies of a single gene or segment of DNA is referred to as cloning DNA.

Cloning vector: DNA molecule originating from a virus, a plasmid, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vectors capacity for self- replication; vectors introduce foreign DNA into host cells, where it can be reproduced in large quantities. Examples are plasmids, cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC); vectors are often recombinant molecules containing DNA sequences from several sources.

cM: See centimorgan.

Code: See codon and genetic code.

Coding sequence: The portion of a gene or an mRNA which actually codes for a protein. Introns are not coding sequences; nor are the 5' or 3' untranslated regions (or the flanking regions, for that matter - they are not even transcribed into mRNA). The coding sequence in a cDNA or mature mRNA includes everything from the AUG (or ATG) initiation codon through to the stop codon, inclusive.

Coding strand: an ambiguous term intended to refer to one specific strand in a double-stranded gene. See "Sense strand".

Codon: A group of three consecutive nucleotides within messenger RNA (mRNA) that encodes a message to initiate translation, to incorporate a specific amino acid into the growing polypeptide chain, or to stop translation. The sequence of codons

in the mRNA unambiguously defines the primary structure of the final protein. Of course, the codons in the mRNA were also present in the genomic DNA, but the sequence may be interrupted by introns. (See genetic code also).

Codon Bias: The tendency for an organism or virus to use certain codons more than others to encode a particular amino acid. An important detrminant 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.

Co-linearity: Refers to an exact correspondence between information encoded in DNA and the polypeptide product ultimately translated from transcripts made from the DNA. Most prokaryotic genes are co-linear with their products; eukaryotic genes often are not.

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.

Complement: Definition determined by context:

1] The complementary sequence to a nucleic acid sequence under study.

2] To provide a function that is missing or altered because of a mutational event.

Complementary DNA (cDNA): A DNA sequence synthesized from a messenger RNA molecule, using reverse transcriptase enzyme. cDNAs can be used experimentally to determine the sequence of messenger RNAs after their introns (non-protein-coding sections) have been spliced out. The single- stranded form is often used as a probe in physical mapping.

Complementary sequences: Nucleic acid base sequences that can form a double- stranded structure by matching base pairs; the complementary sequence to G- T- A- C is C- A- T- G.

Concatemer: Tandem arrays of monomeric DNA molecules with complementary ends. Intra-molecular reassociation of such molecules leads to circularisation while inter-molecular reaction produces concatemers.

Conformation: The 3-dimensional structure of a molecule.

Conformational epitope: Also called a "global epitope". An epitope comprised of contiguous but physically discontinuous components of the immunogenic molecule. In a proteinaceous antigen this could be sequences from different stretches within a polypeptide, or even from different polypeptide subunits.

Conjugation: Physical contact with the establishment of plasma bridges between two different bacterial cells allowing directed transfer of DNA.

Consensus sequence: A term that refers to sequences common to different genes within an organism, or to the same gene among different organisms, that encode a specific function. This term may be applied to either nucleic acids or proteins, since the protein sequence is completely dependent upon the nucleic acid sequence.

Conservation: Identical parts of genes that are present in two distinct organisms are said to be conserved. Conservation can be detected by measuring the similarity of the two sequences at the base (RNA or DNA) or amino-acid (protein) level.

Conserved sequence: A base sequence in a DNA molecule (or an amino acid sequence in a protein) that has remained essentially unchanged throughout evolution. The more similarities there are, the more highly conserved the two sequences.

Consensus sequence: A 'nominal' sequence inferred from multiple, imperfect examples. Multiple lanes of shotgun sequence can me merged to show a consensus sequence. The optimal sequence of nucleotides recognized by some factor. A DNA binding site for a protein may vary substantially, but one can infer the consensus sequence for the binding site by comparing numerous examples. For example, the (fictitious) transcription factor ZQ1 usually binds to the sequences AAAGTT, AAGGTT or AAGATT. The consensus sequence for that factor is said to be AARRTT (where R is any purine, i.e. A or G). ZQ1 may also be able to weakly bind to ACAGTT (which differs by one base from the consensus).

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).

Constitutive expression: Constantly expressed at some appreciable level.

Contig: Several uses, all nouns. The term comes from a shortening of the word 'contiguous'. A 'contig' may refer to a map showing placement of a set of clones that completely, contiguously cover some segment of DNA in which you are interested. Also called the 'minimal tiling path'. More often, the term 'contig' is used to refer to the final product of a shotgun sequencing project. When individual lanes of sequence information are merged to infer the sequence of the larger DNA piece, the product consensus sequence is called a 'contig'.

Contig map: A map depicting the relative order of a linked library of small overlapping clones representing a complete chromosomal segment.

Cosmid: A type of artificially constructed vector used for cloning 35-45 kb of DNA. These are plasmids carrying a phage l cos site (which allows packaging into l capsids), an origin of replication and an antibiotic resistance gene. A plasmid of 40 kb is very difficult to put into bacteria, but can replicate once there. Cosmids, however, have a cos site, and thus can be packaged into l phage heads (a reaction which can be performed *in vitro*) to allow efficient introduction into bacteria.

Cos site:

Crossing over: The breaking during meiosis of one maternal and one paternal chromosome, the exchange of corresponding sections of DNA, and the rejoining of the chromosomes. This process can result in an exchange of alleles between chromosomes. Compare recombination.

Cytosine (C): One of the pyrimidine bases found in DNA and RNA (4-amino-2-hydroxypyrimidine). one member of the base pair G- C (guanine and cytosine).

D

Dalton: Unit of atomic mass. One dalton corresponds to the mass of a hydrogen atom = $3.32X10^{-24}$ g.

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: In molecular biology, this term refers to the fact that multiple different codons may encode the same amino acid. However, a given codon does not encode more than one amino acid within the nucleus of an organism.

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.

Denaturing gradient gel electrophoresis (DGGE): Resolves partially denatured double stranded DNA in precisely conditions of temperature and denaturant concentration. Different alleles may denature to various extents under such conditions, and migrate differently on DGGE acrylamide gels.

Deletion: The absence of bases that are present in the wild-type DNA sequence.

Deoxyribonucleotide: See nucleotide.

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.

Diploid: A full set of genetic material, consisting of paired chromosomes one chromosome from each parental set. Most animal cells except the gametes have a diploid set of chromosomes. The diploid human genome has 46 chromosomes. Compare haploid.

DNA (deoxyribonucleic acid): The molecule responsible for storing and transmitting genetic information. DNA is a double- stranded molecule held together by weak bonds between base pairs of nucleotides twisted around each other in the shape of a double helix. The four nucleotides in DNA contain the bases: adenine (A), guanine (G), cytosine (C), and thymine (T). In nature, base pairs form only between A and T and between G and C; thus the base sequence of each single strand can be deduced from that of its partner.

DNA gyrase: see gyrase.

DNA ligase: Enzymatic activity responsible for creating phosphodiester bonds between the 5' end of one strand of DNA and the 3' end of another strand. Requires the presence of a 5' phosphate on one strand, and a 3' Hydroxyl group on the second strand.

DNA polymerase: Enzymatic activity responsible for catalyzing the polymerization of DNA. Is dependent upon an annealed primer from which to initiate polymerization, and a DNA template from which to copy.

DNA probes: See probe.

DNA replication: The use of existing DNA as a template for the synthesis of new DNA strands. In humans and other eukaryotes, replication occurs in the cell nucleus.

DNase: Deoxyribonuclease, a class of enzymes which digest DNA. The most common is DNase I, an endonuclease which digests both single and double-stranded DNA.

DNA sequence: The relative order of base pairs, whether in a fragment of DNA, a gene, a chromosome, or an entire genome. See base sequence analysis.

Domain: A discrete portion of a protein with its own function. The combination of domains in a single protein determines its overall function.

Dominant: Allele that determines phenotype in a heterozygous individual carrying another recessive allele.

Dot blot: A technique for measuring the amount of one specific DNA or RNA in a complex mixture. The samples are spotted onto a hybridization membrane (such as nitrocellulose or activated nylon, etc.), fixed and hybridized with a radioactive probe. The extent of labeling (as determined by autoradiography and densitometry) is proportional to the concentration of the target molecule in the sample. Standards provide a means of calibrating the results.

Double helix: The helical shape assumed by DNA in which the two complementary strands hydrogen bond together in opposite orientations (i.e. have opposite polarities).

Downstream: See "Upstream/Downstream".

Duplex: A nucleic acid molecule in which two strands are base paired with each other.

E

E. coli (Escherichia coli): A common Gram-negative bacterium present in human intestinal tract that has been studied intensively by geneticists because of its small genome size, normal lack of pathogenicity, and ease of growth in the laboratory. Useful for cloning experiments.

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.

Electrophoresis: A method of separating large molecules (such as DNA fragments or proteins) from a mixture of similar molecules. An electric current is passed through a medium containing the mixture, and each kind of molecule travels through the medium at a different rate, depending on its electrical charge and size. Separation is based on these differences. Agarose and acrylamide gels are the media commonly used for electrophoresis of proteins and nucleic acids.

Elongation factor: A protein(s) that associates with the ribosome cyclically to assist in loading tRNA into the the A site of the ribosome.

End Labeling: The technique of adding a radioactively labeled group to one end (5' or 3' end) of a DNA strand.

Endonuclease: An enzyme that cleaves its nucleic acid substrate at internal sites (as opposed to an exonuclease, which must start at an end) in the nucleotide sequence. Examples include the restriction enzymes, DNase I and RNase A.

Endoplasmic reticulum: A specialized membranous organelle within eukaryotic cells responsible for synthesis of membrane-inserted proteins, and for proteins to be exported of proteins to the cell surface or beyond.

Enhancer: An enhancer is a cis-acting nucleotide sequence to which transcription factor(s) bind, and which increases the transcription of a gene. It is not part of a promoter; the basic difference being that an enhancer can be moved around anywhere in the general vicinity of the gene (within several thousand nucleotides on either side or even within an intron), and it will still function. It can even be clipped out and spliced back in backwards, and will still operate. A promoter, on the other hand, is position- and orientation-dependent. Enhancers are ususlly around 70-80 bp in length and are found, for e.g in viral DNA molecules. Some enhancers are "conditional" - in other words, they enhance transcription only under certain conditions, for example in the presence of a hormone.

Enzyme: A protein that acts as a catalyst, speeding the rate at which a biochemical reaction proceeds but not altering the direction or nature of the reaction.

Epigenetic: A change in phenotype brought about by changes in gene regulation rather than by a change in genotype.

Episome: Extrachromosomal genetic element. Generally used synonymously for plasmid.

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.

EST: see expressed sequence tag and sequence tagged site (STS).

ERE: Estrogen Response Element. A binding site in a promoter to which the activated estrogen receptor can bind. The estrogen receptor is essentially a transcription factor which is activated only in the presence of estrogens. The activated receptor will bind to an ERE, and transcription of the adjacent gene will be altered. See also "Response element".

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.

Euchromatin: The gene-rich regions of a genome (see also heterochromatin).

Eukaryote: Cell or organism with membrane- bound, structurally discrete nucleus and other well- developed subcellular compartments. Eukaryotes include all organisms except viruses, bacteria, and blue- green algae. Compare prokaryote. See chromosomes.

Evolutionary Clock: Defined by the rate at which mutations accumulate within a given gene.

Evolutionarily conserved: See conserved sequence.

Exogenous DNA: DNA originating outside an organism.

Exon: Those portions of a genomic DNA sequence which will be represented in the final, mature mRNA ie. A contiguous segment of genomic DNA that codes for a polypeptide in a gene.The term "exon" can also be used for the equivalent segments in the final RNA. Exons may include coding sequences, the 5' untranslated region or the 3' untranslated region.

Exonuclease: An enzyme that cleaves nucleotides sequentially from free ends of a linear nucleic acid substrate. An example is Exonuclease III, which digests only double-stranded DNA starting from the 3' end.

Expressed sequence tag (EST): A short piece of DNA sequence corresponding to a fragment of a complementary DNA (made from a cell's messenger RNA). ESTs have been used to hunt for genes, so hundreds of thousands are present in sequence databases. (See Sequence tagged sites)

Expression: To "express" a gene is to cause it to function. A gene which encodes a protein will, when expressed, be transcribed and translated to produce that protein. A gene which encodes an RNA rather than a protein (for example, a rRNA gene) will produce that RNA when expressed.

Expression clone: This is a clone (plasmid in a bacteria, or maybe a lambda phage in bacteria) which is designed to produce a protein from the DNA insert. Mammalian genes do not function in bacteria, so to get bacterial expression from your mammalian cDNA, you would place its coding region (i.e. no introns) immediately adjacent to bacterial

transcription/translation control sequences. That artificial construct (the "expression clone") will produce a pseudo-mammalian protein if put back into bacteria. Often, that protein can be recognized by antibodies raised against the authentic mammalian protein, and vice versa.

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.

Expressed gene: See gene expression.

F

Familial: An inherited trait.

FISH (fluorescence in situ hybridization): A physical mapping approach that uses fluorescein tags to detect hybridization of probes with metaphase chromosomes and with the less- condensed somatic interphase chromatin.

Flow cytometry: Analysis of biological material by detection of the light- absorbing or fluorescing properties of cells or subcellular fractions (i.e., chromosomes) passing in a narrow stream through a laser beam. An absorbance or fluorescence profile of the sample is produced. Automated sorting devices, used to fractionate samples, sort successive droplets of the analyzed stream into different fractions depending on the fluorescence emitted by each droplet.

Flow karyotyping: Use of flow cytometry to analyze and/or separate chromosomes on the basis of their DNA content.

Footprinting: A technique by which one identifies a protein binding site on cellular DNA. The presence of a bound protein prevents DNase from "nicking" that region, which can be detected by an appropriately designed gel.

Frame-shift: A change from one reading frame to another.

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.

gDNA: Shorthand for "genomic DNA".

Gel shift assay: (gel mobility shift assay, band shift assay) A method by which one can determine whether a particular protein preparation contains factors which bind to a particular DNA fragment. When a radiolabeled DNA fragment is run on a gel, it shows a characteristic mobility. If it is first incubated with a cellular extract of proteins (or with purified protein), any protein-DNA complexes will migrate slower than the naked DNA - a shifted band.

Gamete: Mature male or female reproductive cell (sperm or ovum) with a haploid set of chromosomes (23 for humans).

Gene: The fundamental physical and functional unit of heredity (Usually DNA, Some organisms have RNA a gene). A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule) that contributes to or influences the phenotype of the cell. A gene usually contains coding regions, introns, untranslated regions and control regions. See gene expression.

Gene Conversion: The alteration of all or part of a gene by a homologous donor DNA that is itself not altered in the process.

Gene expression: The process by which information coded by genes is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

Gene families: Groups of closely related genes that make similar products.

Gene library: See genomic library.

Gene mapping: Determination of the relative positions of genes on a DNA molecule (chromosome or plasmid) and of the distance, in linkage units or physical units, between them.

Gene product: The biochemical material, either RNA or protein, resulting from expression of a gene. The amount of gene product is used to measure how active a gene is; abnormal amounts can be correlated with disease- causing alleles.

Gene superfamily: A large group of genes related (often poorly) by sequence homologies or by the structures of their products, and by their involvement in different aspects of the same larger process (e.g. immunoglobulin gene superfamily).

Gene therapy: A technique involving the use of foreign genetic material to correct a genetic defect or to modify the phenotype of an affected individual, by targeting the somatic cells.

Genetic code: The sequence of nucleotides, coded in triplets (codons) along the mRNA, that determines the sequence of amino acids in protein synthesis. The DNA sequence of a gene can be used to predict the mRNA sequence, and the genetic code can in turn be used to predict the amino acid sequence.

Genetic disease: Heritable genetic alterations from wild type which, when expressed, result in decreased viability of the individual receiving the altered gene(s).

Genetic engineering technologies: See recombinant DNA technologies.

Genetic map: The ordering of genes by the statistical determination of recombination events between them. Genes separated by greater distances are more likely to recombine. (See also linkage map).

Genetic material: See genome.

Genetics: The study of the patterns of inheritance of specific traits.

Genome: The entire compliment of genetic material in the form of permanently maintained DNA for a given organism. Its size is generally given as its total number of base pairs. Mammalian genomic DNA (including that of humans) contains 6 billion base pairs of DNA per diploid cell. There are somewhere in the order of a hundred thousand genes, including coding regions, 5' and 3' untranslated regions, introns, 5' and 3' flanking DNA. Also present in the genome are structural segments such as telomeric and centromeric DNAs and replication origins, and intergenic DNA.

Genome projects: Research and technology development efforts aimed at mapping and sequencing some or all of the genome of human beings and other organisms.

Genomic blot: A type of Southern blot specifically used to analyze a mixture of DNA fragments derived from total genomic DNA. Because genomic DNA is very complicated, when it has been digested with restriction enzymes, it produces a complex set of fragments ranging from tens of bp to tens of thousands of bp. However, any specific gene will be reproducibly found on only one or a few specific

fragments. A million identical cells will produce a million identical restriction fragments for any given gene, so probing a genomic Southern with a gene-specific probe will produce a pattern of perhaps one or just a few bands.

Genomic clone: A piece of DNA taken from the genome of a cell or animal, and spliced into a bacteriophage or other cloning vector. A genomic clone may contain coding regions, exons, introns, 5' flanking regions, 5' untranslated regions, 3' flanking regions, 3' untranslated regions, or it may contain none of these...it may only contain intergenic DNA (usually not a desired outcome of a cloning experiment!).

Genomic library: A collection of clones made from a set of randomly generated overlapping DNA fragments representing the entire genome of an organism. Compare library, arrayed library.

Genotype: The set of genes that an individual carries; usually refers to the particular pair of alleles (alternative forms of a gene) that a person has at a given region of the genome.

Germline: Tissues involved in the generation of haploid gametes.

Glycosylation: The covalent addition of sugar moities to N or O atoms present in the side chains of certain amino acids of certain proteins, generally occuring 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.

GRE: Glucocorticoid Response Element: A binding site in a promoter to which the activated glucocorticoid receptor can bind. The glucocorticoid receptor is essentially a transcription factor which is activated only in the presence of glucocorticoids. The activated receptor will bind to a GRE, and transcription of the adjacent gene will be altered. See also "Response element".

Guanine (G): A nitrogenous base found in DNA and RNA, one member of the base pair G- C (guanine and cytosine). (2-amino-6-hydroxypurine).

Gyrase: An enzymatic activity responsible for maintaining supercoiling in DNA.

H

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

Haploid: A single set of chromosomes (half the full set of genetic material), present in the egg and sperm cells of animals and in the egg and pollen cells of plants. Human beings have 23 chromosomes in their reproductive cells. Compare diploid.

Haplotype: A particular combination of alleles (alternative forms of genes) or sequence variations that are closely linked Ñ that is, are likely to be inherited together Ñ on the same chromosome. Originally was coined to describe MHC antigens, but now is used to describe RFLP patterns and certain other situations.

H-chain: Immunoglobulin heavy chain.

Helix-turn-helix: A protein structural motif characteristic of certain DNA-binding proteins.

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 hydridization 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.

Heterozygous: An individual containing dissimilar alleles for a given gene or locus.

hnRNA (heterogeneous nuclear RNA): Heterogeneous nuclear RNA; refers collectively to the variety of RNAs found in the nucleus, including primary transcripts, partially processed RNAs and snRNA. The term hnRNA is often used just for the unprocessed primary transcripts, however.

Heterochromatin: Compact, gene-poor regions of a genome, which are enriched in simple sequence repeats. As it can be impossible to clone, heterochromatin is often ignored when calculating the percentage of a genome that has been sequenced. Heterochromatin was originally identified as regions of the genome that stained differently to euchromatin (gene-rich regions).

Heterozygosity: The presence of different alleles at one or more loci on homologous chromosomes.

Histones: Highly basic proteins which associate with the chromosomal DNA to package it into a compact, higher order structure.

Homeobox: A short stretch of nucleotides whose base sequence is virtually identical in all the genes that contain it. It has been found in many organisms from fruit flies to human beings. In the fruit fly, a homeobox appears to determine when particular groups of genes are expressed during development.

Homeotic mutation: A mutation in a homeotic gene that is responsible for controlling the activities of numerous other genes, usually during embryologic development in higher organisms.

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.

Homologous chromosomes: A pair of chromosomes containing the same linear gene sequences, each derived from one parent.

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).

Homozygous: An individual containing identical alleles for a given gene or locus.

Host strain (bacterial): The bacterium used to harbor a plasmid. Typical host strains include *HB101* (general purpose *E. coli* strain), *DH5a* (ditto), *JM101* and *JM109* (suitable for growing M13 phages), XL1-Blue (general-purpose, good for blue/white *lacZ* screening). Note that the host strain is available in a form with no plasmids (hence you can put one of your own into it), or it may have plasmids present (especially if you put them there). Hundreds, perhaps thousands, of host strains are available.

Human gene therapy: Insertion of normal DNA directly into cells to correct a genetic defect.

Human Genome Initiative: Collective name for several projects begun in 1986 by Department Of Energy (DOE) to (1) create an ordered set of DNA segments from known chromosomal locations, (2) develop new computational methods for analyzing genetic map and DNA sequence data, and (3) develop new techniques and instruments for detecting and analyzing DNA. This DOE initiative is now known as the Human Genome Program. The national effort, led by DOE and NIH, is known as the Human Genome Project.

Hybridization: The reaction by which the pairing of complementary strands of nucleic acid occurs. DNA is usually double-stranded, and when the strands are separated they will re-hybridize under the appropriate conditions. Hybrids can form between DNA-DNA, DNA-RNA or RNA-RNA. They can form between a short strand and a long strand containing a region complementary to the short one. Imperfect hybrids can also form, but the more imperfect they are, the less stable they will be (and the less likely to form). To "anneal" two strands is the same as to "hybridize" them.

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 hyrophobicity/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.

I

Idiotope: Epitopes present in the unique variable sequences of an immunoglobulin molecule. These may or may not coincide with the immunoglobulin's paratope.

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.

Informatics: The study of the application of computer and statistical techniques to the management of information. In genome projects, informatics includes the development of methods to search databases quickly, to analyze DNA sequence information, and to predict protein sequence and structure from DNA sequence data.

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.

Initiator tRNA: A special tRNA responsible for annealing to the start codon, in the P site of the ribosome, to initiate polypeptide synthesis.

Insert: In a complete plasmid clone, there are two types of DNA - the "vector" sequences and the "insert". The vector sequences are those regions necessary for propagation, antibiotic resistance, and all those mundane functions necessary for useful cloning. In contrast, however, the insert is the piece of DNA in which you are really interested.

Insertion: The presence of additional bases within a sequence that are not present in wild-type sequence.

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.

***In situ* hybridization:** Use of a DNA or RNA probe to detect the presence of the complementary DNA sequence in cloned bacterial or cultured eukaryotic cells.

Intergenic: Between two genes; e.g. intergenic DNA is the DNA found between two genes. The term is often used to mean non-functional DNA (or at least DNA with no known importance to the two genes flanking it). Alternatively, one might speak of the "intergenic distance" between two genes as the number of base pairs from the polyA site of the first gene to the cap site of the second. This usage might therefore include the promoter region of the second gene.

Interphase: The period in the cell cycle when DNA is replicated in the nucleus; followed by mitosis.

Intron: Introns are portions of genomic DNA which are transcribed (and thus present in the primary transcript) but which are later spliced out. They thus are not present in the mature mRNA. Note that although the 3' flanking region is often transcribed, it is removed by endonucleolytic cleavage and not by splicing. It is not an intron. Compare exons.

Inverse PCR: Variation of PCR that makes the amplification of DNA segments of unknown sequence that flank DNA segments of known sequence possible; in brief, total DNA is digested to completion and fragments ligated under conditions that favour circularization of fragments; pair of PCR primers, designed from known sequence known sequence, are used to prime PCR from opposite strands resulting in amplification of fragment of unknown sequence.

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: Literally means "in glass", e.g., test tube; now used to mean growth in any type of culture vessel; Outside a living organism.

In vivo: In living organism

K

Karyotype: A photomicrograph of an individuals chromosomes arranged in a standard format showing the number, size, and shape of each chromosome type; used in low- resolution physical mapping to correlate gross chromosomal abnormalities with the characteristics of specific diseases.

kb:. abbreviation for kilobase, one thousand bases. See kilobase

kd: one thousand Dalton. See Dalton.

Kilobase (kb): Unit of length for DNA fragments equal to 1000 nucleotides.

Kinase: A kinase is in general an enzyme that catalyzes the transfer of a phosphate group from ATP to something else. In molecular biology, it has acquired the more specific verbal usage for the transfer onto DNA of a radiolabeled phosphate group. This would be done in order to use the resultant "hot" DNA as a probe.

Kinteochore: A specialized structure found in the centromeric region of the chromosome that is responsible for attaching to the spindle during nuclear division.

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 experiment: A technique for deleting, mutating or otherwise inactivating a gene in a mouse. This laborious method involves transfecting a crippled gene into cultured embryonic stem cells, searching through the thousands of resulting clones for one in which the crippled gene exactly replaced the normal one (by homologous recombination), and inserting that cell back into a mouse blastocyst. The resulting mouse will be chimaeric but, if you are lucky (and if you've gotten this far, you obviously are), its germ cells will carry the deleted gene. A few rounds of careful breeding can then produce progeny in which both copies of the gene are inactivated.

L-chain: Immunoglobulin light chain.

Leader sequence: Two usages;

1] N-terminal pre sequence of scretory proteins such as peptide hormones and membrane proteins.

2] The untranslated sequence at the 5'-ends of mRNA molecules.

Leucine zipper: A motif found in certain proteins in which Leu residues are evenly spaced through an a-helical region, such that they would end up on the same face of the helix. Dimers can form between two such proteins. The Leu zipper is important in the function of transcription factors such as Fos and Jun and related proteins.

Library: A library might be either a genomic library, or a cDNA library. In either case, the library is just a tube carrying a mixture of thousands of different clones - bacteria or l phages. Each clone carries an "insert" - the cloned DNA.

A cDNA library is usually just a mixture of bacteria, where each bacteria carries a different plasmid. Inserted into the plasmids (one per plasmid) are thousands of different pieces of cDNA (each typ. 500-5000 bp) copied from some source of mRNA, for example, total liver mRNA. The basic idea is that if you have a large enough number of different liver-derived cDNAs carried in those bacteria, there is a 99% probability that a cDNA copy of any given liver mRNA exists somewhere in the tube. The real trick is to find the one you want out of that mess - a process called screening (see "Screening").

A genomic library is similar in concept to a cDNA library, but differs in three major ways - 1) the library carries pieces of genomic DNA (and so contains introns and flanking regions, as well as coding and untranslated); 2) you need bacteriophage l or cosmids, rather than plasmids, because... 3) the inserts are usually 5-15 kb long (in a l library) or 20-40 kb (in a cosmid library). Therefore, a genomic library is most commonly a tube containing a mixture of l phages. Enough different phages must be present in the library so that any given piece of DNA from the source genome has a 99% probability of being present.

Ligase: An enzyme, T4 DNA ligase, which can link pieces of DNA together. The pieces must have compatible ends (both of them blunt, or else mutually compatible sticky ends), and the ligation reaction requires ATP.

Ligation: The process of splicing two pieces of DNA together. In practice, a pool of DNA fragments are treated with ligase (see "Ligase") in the presence of ATP, and all possible splicing products are produced, including circularized forms and end-to-end ligation of 2, 3 or more pieces. Usually, only some of these products are useful, and the investigator must have some way of selecting the desirable ones.

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 measure of proximity of two or more markers (e.g., genes, RFLP markers) on a chromosome determined by recombination events. The closer together the markers are, the lower the probability that they will be separated during DNA repair or replication processes (binary fission in prokaryotes, mitosis or meiosis in eukaryotes), and hence the greater the probability that they will be inherited together.

Linkage map: A map of the relative positions of genetic loci on a chromosome, determined on the basis of how often the loci are inherited together. Distance is measured in centimorgans (cM).

Linker: A small piece of synthetic double-stranded DNA which contains something useful, such as a restriction site. A linker might be ligated onto the end of another piece of DNA to provide a desired restriction site.

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).

Localize: Determination of the original position (locus) of a gene or other marker on a chromosome.

Locus (pl. loci): The position on a chromosome of a gene or other chromosome marker; also, the DNA at that position. The use of locus is sometimes restricted to mean regions of DNA that are expressed. See gene expression.

Long(q) and short(p) arms: The regions either side of the centromere, a compact part of a chromosome, are known as arms. As the centromere is not in the centre of the chromosome, one arm is longer than the other.

M13: A bacteriophage which infects certain strains of *E. coli.* The salient feature of this phage is that it packages only a single strand of DNA into its capsid. If the investigator has inserted some heterologous DNA into the M13 genome, copious quantities of single-stranded DNA can subsequently be isolated from the phage capsids. M13 is often used to generate templates for DNA sequencing.

Macrorestriction map: Map depicting the order of and distance between sites at which restriction enzymes cleave chromosomes.

Maintenance methylase: An enzymatic activity responsible for maintaining the patterns of methylation on each strand of a DNA molecule after replication.

Mapping: See gene mapping, linkage map, physical map.

Marker: Two typical usages:

1] Molecular weight size marker: a piece of DNA of known size, or a mixture of pieces with known size, used on electrophoresis gels to determine the size of unknown DNA's by comparison.

2] Genetic marker: An identifiable physical location on a chromosome (e.g., restriction enzyme cutting site, gene) whose inheritance can be monitored. Markers can be expressed regions of DNA (genes) or some segment of DNA with no known coding function but whose pattern of inheritance can be determined. See RFLP, restriction fragment length polymorphism.

Mb: See megabase.

Megabase (Mb): Unit of length for DNA fragments equal to 1 million nucleotides and roughly equal to 1 cM.

Melting: The dissociation of a duplex nucleic acid molecule into single strands, usually by increasing temperature. See Denaturation.

Meiosis: The process of two consecutive cell divisions in the diploid progenitors of sex cells. Meiosis results in four rather than two daughter cells, each with a haploid set (1n) of chromosomes.

Messenger RNA (mRNA): Proteins are not synthesized directly from genomic DNA. Instead, an RNA template (a precursor mRNA) is constructed from the sequence of the gene. This RNA is then processed in various ways, including splicing. Spliced single-stranded RNAs destined to become templates for protein synthesis are known as mRNAs, The term mRNA is used only for a mature transcript with polyA tail and with all introns removed, rather than the primary transcript in the nucleus. As such, an mRNA will have a 5' untranslated region, a coding region, a 3' untranslated region and (almost always) a poly(A) tail. Typically about 2% of the total cellular RNA is mRNA. See genetic code.

Molecular biology: The biochemical study of the genetic basis for phenotype.

Monocistronic: A form of gene organization resulting in transcription of an mRNA that contains coding sequence for a single gene or gene product.

mRNA: See messenger RNA.

mRNA export: Refers to the movement of spliced mRNA out of the nucleus to the cytoplasm.

mRNA processing: Refers to the processes of polyadenylation, splicing, and addition of a 5' cap structure.

Metaphase: A stage in mitosis or meiosis during which the chromosomes are aligned along the equatorial plane of the cell.

Missense Mutation: A nucleotide mutation which results in a change in the amino acid sequence of the encoded protein (contrast with Silent Mutation).

Mitosis: The process of nuclear division in cells that produces daughter cells that are genetically identical to each other and to the parent cell.

Molecular Genetics: Study of how genes function to control cellular activities.

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.

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.

Multifactorial or multigenic disorders: See polygenic disorders.

Multi-gene family: A group of genes that are related by sequence homologies; usually are also related by their functions or by the processes in which they participate.

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.

Multiplexing: A sequencing approach that uses several pooled samples simultaneously, greatly increasing sequencing speed.

Mutagen: An agent capable of causing mutations. Common examples are ultraviolet light, such as in sunlight, and anthracene, a material formed during the cooking of fatty meats on a barbecue grill.

Mutation: A permanent, heritable change of the genetic material, either in a single gene or in the numbers or structures of the chromosomes. Mutations do not always have harmful effects. Compare polymorphism.

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 amplfication 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 incorporating radioactive isotopes (typically 32P) into a piece of DNA. The DNA is randomly nicked by DNase I, and then starting from those nicks DNA polymerase I digests and then replaces a stretch of DNA. Radiolabeled precursor nucleotide triphosphates can thus be incorporated.

Nitrogenous base: A nitrogen- containing molecule having the chemical properties of a base.

Non-coding strand: Anti-sense strand. See "Sense strand" for a discussion of sense strand vs. anti-sense strand.

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)

Northern blot: A technique for analyzing mixtures of RNA by transfer of size-separated RNA fragments to a synthetic membrane, whereby the presence and rough size of one particular type of RNA (usually an mRNA) can be ascertained. See "Blotting" for more information. After Dr. E. M. Southern invented the Southern blot, it was adapted to RNA and named the "Northern" blot.

Nonsence 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.

nt: Abbreviation for nucleotide; i.e. the monomeric unit from which DNA or RNA are built. One can express the size of a nucleic acid strand in terms of the number of nucleotides in its chain; hence 'nt' can be a measure of chain length.

N Terminus: See Amino Terminus.

Nuclear run-on: A method used to estimate the relative rate of transcription of a given gene, as opposed to the steady-state level of the mRNA transcript (which is influenced not just by transcription rates, but by the stability of the RNA). This technique is based on the assumption that a highly-transcribed gene should have more molecules of RNA polymerase bound to it than will the same gene in a less-active state. If properly prepared, isolated nuclei will continue to transcribe genes and incorporate 32P into RNA, but only in those transcripts that were in progress at the time the nuclei were isolated. Once the polymerase molecules complete the transcript they have in progress, they should not be able to re-initiate transcription. If that is true, then the amount of radiolabel incorporated into a specific type of mRNA is theoretically proportional to the number of RNA polymerase complexes present on that gene at the time of isolation. A very difficult technique, rarely applied appropriately from what I understand.

Nuclease: An enzyme which degrades nucleic acids. A nuclease can be DNA-specific (a DNase), RNA-specific (RNase) or non-specific. It may act only on single stranded nucleic acids, or only on double-stranded nucleic acids, or it may be non-specific with respect to strandedness. A nuclease may degrade only from an end (an exonuclease), or may be able to start in the middle of a strand (an endonuclease). To further complicate matters, many enzymes have multiple functions; for example, Bal31 has a 3'-exonuclease activity on double-stranded DNA, and an endonuclease activity specific for single-stranded DNA or RNA.

Nuclease protection assay: See "RNase protection assay".

Nucleic acid: A large molecule composed of nucleotide subunits.

Nucleoside: A term refering to the combination of adenine, cytosine, guanine, or thymine with a ribose or 2-deoxyribose sugar moiety. A nucleoside is not phosphorylated.

Nucleotide: A building block of DNA or RNA consisting of a nitrogenous base (adenine, guanine, thymine, or cytosine in DNA; adenine, guanine, uracil, or cytosine in RNA), a phosphate molecule, and a sugar molecule (deoxyribose in DNA and ribose in RNA). Thousands of nucleotides are linked to form a DNA or RNA molecule. See DNA, base pair, RNA.

Nucleus: The cellular organelle in eukaryotes that contains the genetic material.

Oligodeoxyribonucleotide: A short, single-stranded DNA molecule, generally l5-50 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: A gene in a tumor virus or in cancerous cells which, when transferred into other cells, can cause transformation (note that only certain cells are susceptible to transformation by any one oncogene). Functional oncogenes are not present in normal cells. A normal cell has many "proto-oncogenes" which serve normal functions, and which under the right circumstances can be activated to become oncogenes. The prefix "v-" indicates that a gene is derived from a virus, and is generally an oncogene (like *v-src , v-ras, v-myb* , etc). See also "Transformation (with respect to cultured cells)".

Oncogenic transformation: A change in the behavioral phenotype of a cell to one lacking in normal proliferative control, and often involving invasive characteristics.

Open reading frame: That segment of a nucleic acid sequence that lies between two stop codons, when translated in a given reading frame. The presence of an

open reading frame is necessary to encode a polypeptide sequence (or an exon thereof), but the presence of an open reading frame is not proof that a polypeptide sequence is encoded in that sequence. See "Reading frame" for a simple example.

Operon: Genes, which are grouped together for coordinate regulation by the same regulator.

Origin of replication: A site at which DNA replication is initiated. There is only one in bacterial chromosomes, but numerous origins in eukaryotic chromosomal DNA. (Abbr. "ori")

Origin recognition complex: An organization of protein factors assembled at an origin of replication for the purpose of initiating replication.

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).

Overlapping clones: See genomic library.

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.

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'

Paratope: That region within the antigen-binding site of an immunoglobulin molecule responsible for recognition of epitope structure.

pBR322: A common plasmid. Along with the obligatory origin of replication, this plasmid has genes which make the *E. coli* host resistant to ampicillin and tetracycline. It also has several restriction sites (BamHI, PstI, EcoRI, HindIII etc.) into which DNA fragments could be spliced in order to clone them.

PCR: See polymerase_chain reaction.

Penetrance: Refers to the proportion of individuals heterozygous for a given dominant allele that express the phenotype of that dominant allele.

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 H2O.

Peptide-binding groove: That region of a Class I or II MHC molecule that is responsible for binding processed antigen peptides for presentation.

Peptidyl transferase: An enzymatic activity of the ribosome responsible for formation of the peptide bond between the nascent polypeptide chain and the amino acid carried by the charged tRNA in the A site of the ribosome. In so doing, the ribosome is moved along the mRNA by one codon.

Phage: A virus for which the natural host is a bacterial cell. Used in the laboratory as a cloning vector. (see bacteriophage)

Phagemid: A type of plasmid which carries within its sequence a bacteriophage replication origin (ori). When the host bacterium is infected with "helper" phage, the phagemid is replicated along with the phage DNA and packaged into phage capsids.

Phase variation: Alternation in the type of flagellum produced by a bacterium.

Phenotype: The observable properties and physical characteristics of of a cell or an organism that is the result of its unique genotype.

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)2 - 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.

Physical map: A map of the locations of identifiable landmarks on DNA (e.g., restriction enzyme cutting sites, genes), regardless of inheritance. Distance is measured in base pairs. For the human genome, the lowest- resolution physical map is the banding patterns on the 24 different chromosomes; the highest-resolution map would be the complete nucleotide sequence of the chromosomes.

Plasmid: A class of circular, autonomously replicating, extrachromosomal DNA elements found in many bacteria. Contain origins of replication to ensure their maintenance. Some plasmids are capable of integrating into the host genome. A number of artificially constructed plasmids are used as cloning vectors or to alter the characteristics of the bacteria. Common plasmids are pBR322, pGEM, pUC18.

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).

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.

PolyA tail: After an mRNA is transcribed from a gene, the cell adds a stretch of A residues (typically 50-200) to its 3' end. It is thought that the presence of this "polyA tail" increases the stability of the mRNA (possibly by protecting it from nucleases). Note that not all mRNAs have a polyA tail; the histone mRNAs in particular do not.

Polycistronic: Refers to a form of gene organization resulting in transcription of an mRNA that contains the coding sequences for multiple gene products, each of which is independently translated from the mRNA.

Polyclonal Antibody: See Monoclonal Antibody.

Polygenic disorders: Genetic disorders resulting from the combined action of alleles of more than one gene (e.g., heart disease, diabetes, and some cancers). Although such disorders are inherited, they depend on the simultaneous presence of several alleles; thus the hereditary patterns are usually more complex than those of single- gene disorders. Compare single- gene disorders.

Polymerase: An enzyme which links individual nucleotides together into a long strand, using another strand as a template. There are two general types of polymerase - DNA polymerases (which synthesize DNA) and RNA polymerase (which makes RNA). Within these two classes, there are numerous sub-types of polymerase, depending on what type of nucleic acid can function as template and what type of nucleic acid is formed. A DNA-dependant DNA polymerase will copy one DNA strand starting from a primer, and the product will be the complementary

DNA strand. A DNA-dependant RNA polymerase will use DNA as a template to synthesize an RNA strand.

Polymerase chain reaction: A technique for replicating a specific piece of DNA *in vitro*, even in the presence of excess non-specific DNA. Primers are added (which initiate the copying of each strand) along with nucleotides and heat stable Taq polymerase. By cycling the temperature, the target DNA is repetitively denatured and copied. Because the newly synthesized DNA strands can subsequently serve as additional templates for the same primer sequences, successive rounds of primer annealing, strand elongation, and dissociation produce rapid and highly specific amplification of the desired sequence. PCR also can be used to detect the existence of the defined sequence in a DNA sample. A single copy of the target DNA, even if mixed in with other undesirable DNA, can be amplified to obtain billions of replicates. PCR can be used to amplify RNA sequences if they are first converted to DNA via reverse transcriptase. This two-phase procedure is known as 'RT-PCR'.

Polymerase Chain Reaction (PCR) is the basis for a number of extremely important methods in molecular biology. It can be used to detect and measure vanishingly small amounts of DNA and to create customized pieces of DNA. It has been applied to clinical diagnosis and therapy, to forensics and to vast numbers of research applications. It would be difficult to overstate the importance of PCR to science.

Polymorphism: Difference in DNA sequence among individuals. To be called a polymorphism, a variant should be present in a significant number of people in the population. Genetic variations occurring in more than 1% of a population would be considered useful polymorphisms for genetic linkage analysis. Compare mutation.

Polynucleotide Kinase: Enzyme which catalyzes the transfer of the terminal phosphate of ATP to 5' hydroxyl termini of polynucleotides, either DNAor RNA. Usually derived from T4 bacteriophage.

Polypeptide: See Peptide.

Polyprotein: A giant polypeptide that contains multiple individual protein sequences embedded within it and which must be proteolytically cleaved to yield the individual proteins.

Polysome: Complex of mRNA and several ribosomes.

Post-transcriptional regulation: Any process occurring after transcription which affects the amount of protein a gene produces. Includes RNA processing efficiency, RNA stability, translation efficiency, protein stability. For example, the rapid degradation of an mRNA will reduce the amount of protein arising from it. Increasing the rate at which an mRNA is translated will increase the amount of protein product.

Post-translational processing: The reactions which alter a protein's covalent structure, such as phosphorylation, glycosylation or proteolytic cleavage.

Post-translational regulation: Any process which affects the amount of protein produced from a gene, and which occurs AFTER translation in the grand scheme of genetic expression. Actually, this is often just a buzz-word for regulation of the stability of the protein. The more stable a protein is, the more it will accumulate.

Primase: An DNA-dependent RNA polymerase function that synthesizes a short RNA primer used during DNA replication by DNA polymerase. Primase does not require a primer, only a template.

PRE: Progesterone Response Element: A binding site in a promoter to which the activated progesterone receptor can bind. The progesterone receptor is essentially a transcription factor which is activated only in the presence of progesterone . The activated receptor will bind to a PRE, and transcription of the adjacent gene will be altered. See also "Response element".

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).

Primary transcript: When a gene is transcribed in the nucleus, the initial product is the primary transcript, an RNA containing copies of all exons and introns. This primary transcript is then processed by the cell to remove the introns, to cleave off unwanted 3' sequence, and to polyadenylate the 5' end. The mature message thus formed is then exported to the cytoplasm for translation.

Primer: A small oligonucleotide (anywhere from 6 to 50 nt long) used to prime DNA synthesis. The DNA polymerases are only able to extend a pre-existing strand along a template; they are not able to take a naked single strand and produce a complementary copy of it de-novo. A primer which sticks to the template is therefore used to initiate the replication. Primers are necessary for DNA sequencing and PCR.

Primer extension: This is a method used to figure out how far upstream from a fixed site the start of an mRNA is. For example, perhaps you have isolated a cDNA clone, but you don't think that the clone has all of the 5' untranslated region. To find out how much is missing, you would first sequence the part you have, and figure out which strand is coding strand (usually the coding strand will have a large open reading frame). Next, you ask the DNA Synthesis Facility to make an oligonucleotide complementary to the 5'-most region of the coding strand (and thus complementary to the mRNA). This "primer" is hybridized to mRNA (say, a mixture

of mRNA containing the one in which you are interested), and reverse transcriptase is added to copy the mRNA from the primer out to the 5' end. The size of the resulting DNA fragment shows how far away from the 5' end your primer is.

Prion: An infectious agent thought to be composed solely of protein. Term is derived from "Protein infectious agent".

Prokaryote: A single-celled organism with a simple internal structure and no nucleus. Bacteria and archaebacteria are prokaryotes.

Promoter site: Region of a DNA molecule 5' to a coding sequence that is responsible for assembly of an RNA polymerase complex and the initiation of transcription.

Proof-reading: Mechanism for correction of errors made during synthesis of nucleic acids or polypeptides by scrutiny of the products after the nucleotides or amino acids have already been incorporated.

Proteasome: A specialized organelle within the cytoplasm of a cell that is responsible for degradation of cytoplasmically situated proteins. The proteasome plays a key role in normal protein turnover and in peptide presentation by MHC Class I antigen.

Protein: The complete, assembled form of a holoprotein, containing all necessary subunits (eg. hemoglobin is comprised of two a and two b globin subunits, as well as a heme prosthetic group) (see also polypeptide).

Proteome: The complete set of proteins encoded by the genome.

Probe: Single- stranded DNA or RNA fragment of specific base sequence, labeled either radioactively (often incorporating 32P or 35S) or immunologically, that are used to detect the complementary base sequence by hybridization. For example, if you want to quantitate the levels of alpha subunit mRNA in a preparation of pituitary RNA, you might make a radiolabeled RNA in-vitro which is complementary to the mRNA, and then use it to probe a Northern blot of the pit RNA. A probe can be radiolabeled, or tagged with another functional group such as biotin. A probe can be cloned DNA, or might be a synthetic DNA strand. As an example of the latter, perhaps you have isolated a protein for which you wish to obtain a cDNA or genomic clone. You might (pay to) microsequence a portion of the protein, deduce the nucleic acid sequence, (pay to) synthesize an oligonucleotide carrying that sequence, radiolabel it and use it as a probe to screen a cDNA library or genomic library. A better way is to call up someone who already has the clone.

Prokaryote: Cell or organism lacking a membrane- bound, structurally discrete nucleus and other subcellular compartments. Bacteria are prokaryotes. Compare eukaryote. See chromosomes.

Processing: The reactions occurring in the nucleus which convert the primary RNA transcript to a mature mRNA. Processing reactions include capping, splicing and polyadenylation. The term can also refer to the processing of the protein product, including proteolytic cleavages, glycosylation, etc.

Promoter: The first few hundred nucleotides of DNA "upstream" (on the 5' side) of a gene, which control the transcription of that gene. The promoter is part of the 5' flanking DNA, i.e. it is not transcribed into RNA, but without the promoter, the gene is not functional. Note that the definition is a bit hazy as far as the size of the region encompassed, but the "promoter" of a gene starts with the nucleotide immediately upstream from the cap site, and includes binding sites for one or more transcription factors which can not work if moved farther away from the gene.

Protein: A large molecule composed of one or more chains of amino acids in a specific order; the order is determined by the base sequence of nucleotides in the gene coding for the protein. Proteins are required for the structure, function, and regulation of the bodys cells, tissues, and organs, and each protein has unique functions. Examples are hormones, enzymes, and antibodies.

Proto-oncogene: A gene present in a normal cell which carries out a normal cellular function, but which can become an oncogene under certain circumstances. The prefix "c-" indicates a cellular gene, and is generally used for proto-oncogenes (examples: *c-myb , c-myc , c-fos , c-jun* , etc).

Pseudogene: A region of DNA that shows extensive similarity to a known gene, but which cannot itself function, either because it has lost the signal required for transcription (the promoter sequence) or because it carries mutations that prevent it from being translated into protein. It is generally assumed that pseudogenes are copies of mRNA. Pseudogenes are known in the globin system but also in many other multigene families.

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.

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.

Pulsed field gel electrophoresis (PFGE): A gel technique which allows size-separation of very large fragments of DNA, in the range of hundreds of kb to thousands of kb. As in other gel electrophoresis techniques, populations of

molecules migrate through the gel at a speed related to their size, producing discrete bands. In normal electrophoresis, DNA fragments greater than a certain size limit all migrate at the same rate through the gel. In PFGE, the electrophoretic voltage is applied alternately along two perpendicular axes, which forces even the larger DNA fragments to separate by size.

Purine: A nitrogen- containing, single- ring, basic compound that occurs in nucleic acids. The purines in DNA and RNA are adenine and guanine.

Pyrimidine: A nitrogen- containing, double- ring, basic compound that occurs in nucleic acids. The pyrimidines in DNA are cytosine and thymine; in RNA, cytosine and uracil.

R

Random primed synthesis: If you have a DNA clone and you want to produce radioactive copies of it, one way is to denature it (separate the strands), then hybridize to that template a mixture of all possible 6-mer oligonucleotides. Those oligos will act as primers for the synthesis of labeled strands by DNA polymerase (in the presence of radiolabeled precursors).

Rare- cutter enzyme: See restriction enzyme cutting site.

Reading frame: When mRNA is translated by the cell, the nucleotides are read three at a time. By starting at different positions, the groupings of three that are produced can be entirely different. The following example shows a DNA sequence and the three reading frames in which it could be read. Not only is an entirely different amino acid sequence specified by the different reading frames, but two of the three frames have stop codons, and thus are not open reading frames (asterisks indicate a stop codon).

A DNA open reading frame: ...ATG ACA TGT AAA GAT AGA CTA ACC TTT TGG...

...Met Thr Cys Lys Asp Arg Leu Thr Phe Trp...

Same bases, different grouping: ...A TGA CAT GTA AAG ATA GAC TAA CCT TTT GG...

... *** His Val Lys Ile Asp *** Pro Phe Gly..

Same bases, another grouping: ...AT GAC ATG TAA AGA TAG ACT AAC CTT TTG G..

... Asp Met *** Arg *** Thr Asn Leu Leu ...

If we shift the grouping again, we will just get the first reading frame again. The reading frame that is actually used is determined by the first methionine codon (the initiation codon). Once that first AUG is recognized, the pattern of triplet groupings follows unambiguously.

Recessive: Allele that determines phenotype only when homozygous; does not affect phenotype when heterozygous with a dominant allele.

Recombinant clones: Clones containing recombinant DNA molecules. See recombinant DNA technologies.

Recombinant DNA: The combination of foreign DNA inserts with vector DNA (e.g., plasmid, phage, cosmid, etc.) to produce a clone within a host.

Recombinant DNA technologies: Procedures used to join together DNA segments in a cell- free system (an environment outside a cell or organism). Under appropriate conditions, a recombinant DNA molecule can enter a cell and replicate there, either autonomously or after it has become integrated into a cellular chromosome.

Recombination: The process by which DNA is exchanged between pairs of equivalent chromosomes (crossing over) during egg and sperm formation. Recombination has the effect of making the chromosomes of the offspring distinct from those of the parents.

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.

Recombinase: Enzymatic activity responsible for facilitating intra- or intermolecular recombination of DNA molecules.

Regulatory regions or sequences: A DNA base sequence that controls gene expression.

Relaxed DNA: See Supercoil.

Repetitive DNA: A surprising portion of any genome consists not of genes or structural elements, but of frequently repeated simple sequences. These may be short repeats just a few nt long, like CACACA etc. They can also range up to a few hundred nt long. Examples of the latter include Alu repeats, LINEs, SINEs. The function of these elements is often unknown. In shorter repeats like di- and tri-nucleotide repeats, the number of repeating units can occasionally change during evolution and descent. They are thus useful markers for familial relationships and

have been used in paternity testing, forensic science and in the identification of human remains.

Replication: The act of a cell making a copy of all or some part its genomic DNA.

Replicon: A segment of genomic DNA that contains an origin of replication and is replicated under the control of that origin.

Reporter Gene: The use of a functional enzyme, such as beta-galactosidase, luciferase, or chloramphenicol acetyltransderase, 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: A form of gene regulation wherein the promoter is prevented from assembling an RNA polymerase complex, so that transcription does not occur.

Resolution: Degree of molecular detail on a physical map of DNA, ranging from low to high.

Response element: By definition, a "response element" is a portion of a gene which must be present in order for that gene to respond to some hormone or other stimulus. Response elements are binding sites for transcription factors. Certain transcription factors are activated by stimuli such as hormones or heat shock. A gene may respond to the presence of that hormone because the gene has in its promoter region a binding site for hormone-activated transcription factor. Example: the glucocorticoid response element (GRE).

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: To "restrict" DNA means to cut it with a restriction enzyme. See "Restriction Enzyme".

Restriction enzymes : An endonuclease enzyme, isolated from bacteria, that recognizes specific base-pair sequences within DNA and causes endonucleolytic cleavage of the DNA at a site determined by the recognized DNA sequences. The sites vary from frequent to rare cutting, depending upon the length of the restriction site.

For example, the restriction enzyme BamHI locates and cuts any occurrence of:

```
5'-GGATCC-3'

||||||
```

3'-CCTAGG-5'

Note that both strands contain the sequence GGATCC, but in antiparallel orientation. The recognition site is thus said to be palindromic, which is typical of restriction sites. Every copy of a plasmid is identical in sequence, so if BamHI cuts a particular circular plasmid at three sites producing three "restriction fragments", then a million copies of that plasmid will produce those same restriction fragments a million times over.

Bacteria produce restriction enzymes for protection against invasion by foreign DNA such as phages. The bacteria's own DNA is modified in such a way as to prevent it from being clipped. Bacteria contain over 600 such enzymes that recognize and cut over 100 different DNA sequences. See restriction enzyme cutting site.

Restriction enzyme cutting site: A specific nucleotide sequence of DNA at which a particular restriction enzyme cuts the DNA. Some sites occur frequently in DNA (e.g., every several hundred base pairs), others much less frequently (rare-cutter; e.g., every 10,000 base pairs).

Restriction fragment: The piece of DNA released after restriction digestion of plasmids or genomic DNA. See "Restriction enzyme". One can digest a plasmid and isolate one particular restriction fragment (actually a set of identical fragments). The term also describes the fragments detected on a genomic blot which carry the gene of interest.

Restriction map: A "cartoon" depiction of the locations within a stretch of known DNA where restriction enzymes will cut.

The map usually indicates the approximate length of the entire piece (scale on the bottom), as well as the position within the piece at which designated enzymes will cut. This map happens to be of a plasmid, and the two ends are joined together with about 25 nt between the EcoRI and HindIII sites.

Restriction site: See Restriction enzyme.

Restriction fragment length polymorphism (RFLP): Variation in the distance between restriction enzyme cleavage sites that exist within a population producing unique DNA fingerprint patterns.

Restriction fragment length polymorphism (RFLP): Variation between individuals in DNA fragment sizes cut by specific restriction enzymes; polymorphic sequences that result in RFLPs are used as markers on both physical maps and genetic linkage maps. RFLPs are usually caused by mutation at a cutting site. See marker.

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: Singel stranded RNA virus which is replicated and expressed via a double stranded DNA intermediate. Retroviruses contain enzyme reverse transcriptase (RNA dependent DNA polymerase); this enzyme converts viral RNA into DNA which can combine with DNA of host cell and produce more viral particles.

Revertant: See Back Mutation.

RFLP: Restriction fragment length polymorphism; the acronym is pronounced "riflip". Although two individuals of the same species have almost identical genomes, they will always differ at a few nucleotides. Some of these differences will produce new restriction sites (or remove them), and thus the banding pattern seen on a genomic Southern will thus be affected. For any given probe (or gene), it is often possible to test different restriction enzymes until you find one which gives a pattern difference between two individuals - a RFLP. The less related the individuals, the more divergent their DNA sequences are and the more likely you are to find a RFLP.

RFLP: See restriction fragment length polymorphism.

Reverse transcriptase: An enzyme found in the retroviruses which catalyzes the RNA-dependent polymerization of DNA (DNA copy from RNA template). RT is used to make cDNA; one begins by isolating polyadenylated mRNA, providing oligo-dT as a primer, and adding nucleotide triphosphates and RT to copy the RNA into cDNA.

Ribonuclease: see "RNAse".

Ribonucleic acid (RNA): A chemical found in the nucleus and cytoplasm of cells; it plays an important role in protein synthesis and other chemical activities of the cell. The structure of RNA is similar to that of DNA. There are several classes of RNA molecules, including messenger RNA, transfer RNA, ribosomal RNA, and other small RNAs, each serving a different purpose.

Ribonucleotides: See nucleotide.

Riboprobe: A strand of RNA synthesized in-vitro (usually radiolabeled) and used as a probe for hybridization reactions. An RNA probe can be synthesized at very high specific activity, is single stranded (and therefore will not self anneal), and can be used for very sensitive detection of DNA or RNA.

Ribosome: The large, multi-subunit ribonucleoprotein cellular complex responsible for translation of mRNA into polypeptide sequences. Ribosomes are a complex consisting of ribosomal RNAs (rRNA) and several proteins.

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 binding site: Sequences contained in an mRNA that organize the assembly of a ribosome to initiate translation of the mRNA into polypeptide.

Ribosomal RNA (rRNA): A class of RNA found in the ribosomes of cells.

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: See ribonucleic acid.

RNase: Ribonuclease; an enzyme which degrades RNA. It is ubiquitous in living organisms and is exceptionally stable. The prevention of RNase activity is the primary problem in handling RNA. There are many different RNases, some of the more important include:

RNase A Cleaves ssRNA 3' of pyrimidines

RNase T1 Cleaves ssRNA at guanasine nucleotides

RNase V1 Cleaves dsRNA (helical regions)

RNase H Degrades the RNA part of RNA:DNA hybrids.

RNA polymerase: Enzymatic activity responsible for DNA-dependent synthesis of RNA. In prokaryotes there is only one RNA polymerase. In eukaryotes, there are three, each of which transcribes a different group of genes.

RNase protection assay: This is a sensitive method to determine (1) the amount of a specific mRNA present in a complex mixture of mRNA and/or (2) the sizes of

exons which comprise the mRNA of interest. A radioactive DNA or RNA probe (in excess) is allowed to hybridize with a sample of mRNA (for example, total mRNA isolated from tissue), after which the mixture is digested with single-strand specific nuclease. Only the probe which is hybridized to the specific mRNA will escape the nuclease treatment, and can be detected on a gel. The amount of radioactivity which was protected from nuclease is proportional to the amount of mRNA to which it hybridized. If the probe included both intron and exons, only the exons will be protected from nuclease and their sizes can be ascertained on the gel.

RNA Splicing: A complex and incompletly understood series of reactions occuring 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.

rRNA: "ribosomal RNA"; any of several RNAs which become part of the ribosome, and thus are involved in translating mRNA and synthesizing proteins. They are the most abundant RNA in the cell (on a mass basis).

RT-PCR: See 'Polymerase Chain Reaction and reverse transcriptase'.

S1 end mapping: A technique to determine where the end of an RNA transcript lies with respect to its template DNA (the gene). Can't be described in a short paragraph. See "RNAse Protection assay" for a closely related technique.

S1 nuclease: An enzyme which digests only single-stranded nucleic acids.

Screening: To screen a library (see "Library") is to select and isolate individual clones out of the mixture of clones. For example, if you needed a cDNA clone of the pituitary glycoprotein hormone alpha subunit, you would need to make (or buy) a pituitary cDNA library, then screen that library in order to detect and isolate those few bacteria carrying alpha subunit cDNA.

There are two methods of screening which are particularly worth describing: screening by hybridization, and screening by antibody.

Screening by hybridization involves spreading the mixture of bacteria out on a dozen or so agar plates to grow several ten thousand isolated colonies. Membranes are laid onto each plate, and some of the bacteria from each colony stick, producing replicas of each colony in their original growth position). The membranes are lifted and the adherent bacteria are lysed, then hybridized to a radioactive piece of alpha DNA (the source of which is a story in itself - see "Probe"). When X-ray film is laid on the filter, only colonies carrying alpha sequences will "light up". Their position on the membranes show where they grew on the original plates, so you now can go back to the original plate (where the

remnants of the colonies are still alive), pick the colony off the plate and grow it up. You now have an unlimited source of alpha cDNA.

Screening by antibody is an option if the bacteria and plasmid are designed to express proteins from the cDNA inserts (see "Expression clones"). The principle is similar to hybridization, in that you lift replica filters from bacterial plates, but then you use the antibody (perhaps generated after olde tyme protein purification rituals) to show which colony expresses the desired protein.

SDS: Sodium dodecyl sulphate. It is an ionic detergent.

SDS-PAGE: 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.

Sense strand: A gene has two strands: the sense strand and the anti-sense strand. The Sense strand is, by definition, the same 'sense' as the mRNA; that is it can be translated exactly as the mRNA sequence can. Given a sense strand with the following sequence:

```
5' - ATG GGG CCA CGG CTG TGA - 3'

Met Gly Pro Arg Leu stop
```

The anti-sense strand will read as follows (note that the strand has been reversed and complemented):

5' - TCA CAG CCG TGG CCC CAT - 3'

The duplex DNA will pair as follows:

```
5' - ATGGGGCCACGGCTGTGA - 3'

||||||||||||||||||
```

3' - TACCCCGGAGCCGACACT - 5'

Note however that when the RNA is transcribed from this sequence, the ANTI-SENSE strand is used as the template for RNA polymerization. After all, the RNA must base-pair with its template strand (see Figure 3), so the process of transcription produces the complement of the anti-sense strand. This introduces some confusion about terminology:

Some people use the term 'coding strand' and 'non-coding strand' to refer to the sense and antisense strands, respectively. Unfortunately, many people interpret these terms in exactly the opposite way. I consider the terms 'coding strand' and 'non-coding strand' to be too ambiguous. Some people use the exact opposite definition for 'sense' and 'anti-sense' that I have given here. Be aware of the possibility of a discrepancy. Textbooks I have consulted generally agree with the nomenclature given herein, albeit some avoid defining these terms at all.

Sequence: As a noun, the sequence of a DNA is a buzz word for the structure of a DNA molecule, in terms of the sequence of bases it contains. As a verb, "to sequence" is to determine the structure of a piece of DNA; i.e. the sequence of nucleotides it contains.

Sequence homology: Similarities in nucleic acid sequence and organization, and in their encoded products, that are sufficiently great as to imply common ancestral origins.

Sequence Polymorphism: See Polymorphism.

Sequence similarity: Similarity in nucleic acid or polypeptide sequences, particularly in shorter segments, that may not be sufficient to imply common ancestral origins.

Sequence tagged site (STS): Short (200 to 500 base pairs) DNA sequence that has a single occurrence in the human genome and whose location and base sequence are known. Detectable by polymerase chain reaction, STSs are useful for localizing and orienting the mapping and sequence data reported from many different laboratories and serve as landmarks on the developing physical map of the human genome. Expressed sequence tags (ESTs) are STSs derived from cDNAs. (See EST)

Sequential Epitope: See Linear Epitope.

Sequencing: Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule or the order of amino acids in a protein.

Sex chromosomes: The X and Y chromosomes in human beings that determine the sex of an individual. Females have two X chromosomes in diploid cells; males have an X and a Y chromosome. The sex chromosomes comprise the 23rd chromosome pair in a karyotype. Compare autosome.

Sex-linked: A trait transmitted by one of the chromosomes determining sex of an individual.

Short arm (p): One of the two prominent segments of a chromosome; the long or "q" arm is the other. The arms of a given chromosome join at its centromere.

Shotgun cloning: The practice of randomly clipping a larger DNA fragment into various smaller pieces, cloning everything, and then studying the resulting individual clones to figure out what happened. For example, if one was studying a 50 kb gene, it "may" be a bit difficult to figure out the restriction map. By randomly breaking it into smaller fragments and mapping those, a master restriction map could be deduced. See also Shotgun sequencing.

Shotgun sequencing: A way of determining the sequence of a large DNA fragment which requires little brain power but lots of late nights. The large fragment is shotgun cloned (see above), and then each of the resulting smaller clones ("subclones") is sequenced. By finding out where the subclones overlap, the sequence of the larger piece becomes apparent. Note that some of the regions will get sequenced several times just by chance.

Shuttle vector: A type of cloning vector that contains sequences enabling it to be propagated and maintained in more than one type of host (e.g. *E. coli* and mammalian cells). For this purpose shuttle vectors carry different origin of DNA replication which are characteristic of the desired host systems.

Sigma Factor: Subunit of bacterial RNA polymerase which controls the correct initiation of transcription. These proteins increase the binding affinity of RNA polymerase to a promoter. Different sigma factors recognize different promoter sequences.

Signal recognition particle (SRP): A chaperonin complex responsible for arresting polypeptide synthesis and facilitating the docking of a ribosome to the endoplasmic reticulum membrane. Normally, on ribosomes translating polypeptides destined for insertion into or across the endoplasmic reticulum membrane become associated with SRP.

Signal Peptidase: An enzyme present within the lumen of the endoplasmic reticulum which proteolytically cleaves a secreted protein at the site of a signal sequence.

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.

Single- gene disorder: Hereditary disorder caused by a mutant allele of a single gene (e.g., Duchenne muscular dystrophy, retinoblastoma, sickle cell disease). Compare polygenic disorders.

Single nucleotide polymorphism (SNP): A polymorphism caused by the change of a single nucleotide. Most genetic variation between individual humans is believed to be due to SNPs.

Single-strand conformational polymorphism (SSCP): Relies on secondary and tertiary structural differences between denatured and rapidly cooled amplified DNA fragments that differ slightly in their DNA sequence; different SSCP alleles are resolved on non-denaturing acrylamide gels, usually at low temperature; ability to resolve alleles depends on conditions of electrophoresis.

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.

Slot blot: Similar to a dot blot, but the analyte is put onto the membrane using a slot-shaped template. The template produces a consistently shaped spot, thus decreasing errors and improving the accuracy of the analysis. See Dot blot.

snRNA: Small nuclear RNA; forms complexes with proteins to form snRNPs; involved in RNA splicing, polyadenylation reactions, other unknown functions (probably).

snRNP: "snerps", Small Nuclear RiboNucleoProtein particles, which are complexes between small nuclear RNAs and 7-10 proteins, and which are involved in RNA splicing and polyadenylation reactions.

Solution hybridization: A method closely related to RNase protection (see "Rnase protection assay"). Solution hybridization is designed to measure the levels of a specific mRNA species in a complex population of RNA. An excess of

radioactive probe is allowed to hybridize to the RNA, then single-strand specific nuclease is used to destroy the remaining unhybridized probe and RNA. The "protected" probe is separated from the degraded fragments, and the amount of radioactivity in it is proportional to the amount of mRNA in the sample which was capable of hybridization. This can be a very sensitive detection method.

Somatic: Refers to non-germline cells. Somatic cells may become terminally differentiated with alterations in their overall genetic compliment, because they are not responsible for passing along the organism's genetic material to the offspring.

Somatic cells: Any cell in the body except gametes and their precursors.

Somatic hypermutation: A very high frequency of mutational events that occur in specific loci, such as the variable segments of expressed immunoglobulin genes. Somatic hypermutation in immunoglobulin genes occurs after all rearrangements have occurred and provide additional potential for variation in immunoglobulin structure.

Southern blot: The transfer (by absorption) of size-separated (by electrophoresis) DNA fragments to a synthetic membrane whereby the presence and rough size of one particular fragment of DNA can be ascertained by detection of specific base sequences ascertained by radiolabeled complementary probes. Initially described by E.M. Southern.

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.

Splice sequence: A sequence within an mRNA molecule that is the site at which splicing occurs.

Splicing: The process that removes introns (non-protein-coding portions) from transcribed RNAs. Exons (protein-coding portions) can also be removed. Depending on which exons are removed, different proteins can be made from the same initial RNA or gene. Different proteins created in this way are 'splice variants' or 'alternatively spliced'.

Splicosome: A ribonucleoprotein structure responsible for splicing of primary transcripts.

Stable transfection: A form of transfection experiment designed to produce permanent lines of cultured cells with a new gene inserted into their genome.

Usually this is done by linking the desired gene with a "selectable" gene, i.e. a gene which confers resistance to a toxin (like G418, aka Geneticin). Upon putting the toxin into the culture medium, only those cells which incorporate the resistance gene will survive, and essentially all of those will also have incorporated the experimenter's gene.

Start codon: That codon at which translation of an mRNA molecule begins. This is always an AUG (encoding methionine) in eukaryotes, and nearly always in prokaryotes. In prokaryotes N-formyl-methionine is used to initiate polypeptide synthesis.

Sticky ends: After digestion of a DNA with certain restriction enzymes, the ends left have one strand overhanging the other to form a short (typically 4 nt) single-stranded segment. This overhang will easily re-attach to other ends like it, and are thus known as "sticky ends". For example, the enzyme BamHI recognizes the sequence GGATCC, and clips after the first G in each strand:

The overhangs thus produced can still hybridize ("anneal") with each other, even if they came from different parent DNA molecules, and the enzyme ligase will then covalently link the strands. Sticky ends therefore facilitate the ligation of diverse segments of DNA, and allow the formation of novel DNA constructs.

Stop codon: That codon at which translation of an mRNA molecule into a polypeptide is terminated. In the Universal Code this may be: UGA, UAG, or UAA.

Streptavidin: A bacterial analog of egg white avidin.

Stringency: A term used to describe the conditions of hybridization. By varying the conditions (especially salt concentration and temperature) a given probe sequence may be allowed to hybridize only with its exact complement (high stringency), or with any somewhat related sequences (relaxed or low stringency). Increasing the temperature or decreasing the salt concentration will tend to increase the selectivity of a hybridization reaction, and thus will raise the stringency.

STS: See sequence tagged site.

Sub-cloning: If you have a cloned piece of DNA (say, inserted into a plasmid) and you need unlimited copies of only a part of it, you might "sub-clone" it. This involves starting with several million copies of the original plasmid, cutting with restriction enzymes, and purifying the desired fragment out of the mixture. That fragment can then be inserted into a new plasmid for replication. It has now been subcloned.

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.

Tandem repeat sequences: Multiple copies of the same base sequence on a chromosome; used as a marker in physical mapping.

Taq polymerase: A DNA polymerase isolated from the bacterium *Thermophilis aquaticus* and which is very stable to high temperatures. It is used in PCR procedures and high temperature sequencing.

TATA box: A sequence found in the promoter (part of the 5' flanking region) of many genes. Deletion of this site (the binding site of transcription factor TFIID) causes a marked reduction in transcription, and gives rise to heterogeneous transcription initiation sites.

Technology transfer: The process of converting scientific findings from research laboratories into useful products by the commercial sector.

Telomerase: A ribonucleoprotein complex that maintains the repeat sequence structures at the telomeric ends of chromosomes.

Telomere: The natural distal end of a chromosome. Contain some form of simple repeating sequence, usually with a single stranded distal end that may form a hairpin. These specialized structures are involved in the replication and stability of linear DNA molecules. See DNA replication.

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.

Tet resistance: See "Antibiotic resistance".

Thymine (T): One of the pyrimidine bases found in DNA one member of the base pair A- T (adenine- thymine). (2,4-dihydroxy-5-methylpyrimidine).

Tissue-specific expression: Gene function which is restricted to a particular tissue or cell type. For example, the glycoprotein hormone alpha subunit is produced only in certain cell types of the anterior pituitary and placenta, not in lungs or skin; thus expression of the glycoprotein hormone alpha-chain gene is said to be tissue-specific. Tissue specific expression is usually the result of an enhancer which is activated only in the proper cell type.

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.

Topoisomerase: An enzymatic activity responsible for relieving excessive supercoiling in DNA.

Trans-: Located on two physically dis-contiguous DNA molecules.

Transcription: The process of copying a gene into RNA transcript. This is the first step in the expression of any gene. The resulting RNA, if it codes for a protein, will be spliced, polyadenylated, transported to the cytoplasm, and by the process of translation will produce the desired protein molecule, although not all transcripts lead to proteins. Compare translation.

Transcription factor: A protein which controls the transcription of genes. These usually bind to DNA as part of their function (but not necessarily). A transcription factor may be general (i.e. acting on many or all genes in all tissues), or tissue-specific (i.e. present only in a particular cell type, and activating the genes restricted to that cell type). Its activity may be constitutive, or may depend on the presence of some stimulus; for example, the glucocorticoid receptor is a transcription factor which is active only when glucocorticoids are present.

Transcription start site: The point in a DNA sequence at which transcription of a gene into RNA begins.

Transcriptome: The complete set of RNAs transcribed from a genome.

Transduction: The incorporation of a cellular gene into a viral genome, that can then be introduced into other cells.

Transfection: A method by which foreign DNA may be put into a cultured mammalian cell. Such experiments are usually performed using cloned DNA containing coding sequences and control regions (promoters, etc) in order to test

whether the DNA will be expressed. Since the cloned DNA may have been extensively modified (for example, protein binding sites on the promoter may have been altered or removed), this procedure is often used to test whether a particular modification affects the function of a gene.

Transfer RNA (tRNA): A class of RNA having structures with triplet nucleotide sequences that are complementary to the triplet nucleotide coding sequences of mRNA. The role of tRNAs in protein synthesis is to bond with amino acids and transfer them to the ribosomes, where proteins are assembled according to the genetic code carried by mRNA.

Transformation: A process by which the genetic material carried by an individual cell is altered by incorporation of exogenous DNA into its genome. The cancerous alteration of mammalian cells

Transformation (with respect to cultured cells): A change in cell morphology and behavior which is generally related to carcinogenesis. Transformed cells tend to exhibit characteristics known collectively as the "transformed phenotype" (rounded cell bodies, reduced attachment dependence, increased growth rate, loss of contact inhibition, etc). There are different "degrees" of transformation, and cells may exhibit only a subset of these characteristics. Not well understood, the process of transformation is the subject of intense research.

Transformed phenotype: Acquisition of a phenotype characteristic of oncogenic transformation.

Transgenic: Creation of an individual in which genetic modification has occured in the germline tissues and may be transmitted to subsequent generations.

Transgenic mouse: A mouse which carries experimentally introduced DNA. The

procedure by which one makes a transgenic mouse involves the injection of DNA into a fertilized embryo at the pro-nuclear stage. The DNA is generally cloned, and may be experimentally altered. It will become incorporated into the genome of the embryo. That embryo is implanted into a foster mother, who gives birth to an animal carrying the new gene. Various experiments are then carried out to test the functionality of the inserted DNA.

Transient transfection: When DNA is transfected into cultured cells, it is able to stay in those cells for about 2-3 days, but then will be lost (unless steps are taken to ensure that it is retained - see Stable transfection). During those 2-3 days, the DNA is functional, and any functional genes it contains will be expressed. Investigators take advantage of this transient expression period to test gene function.

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 of converting the genetic code into polypeptides, catalyzed by the ribosome and a host of soluble factors. mRNA codons are recognized by tRNA anti-codons. Each tRNA codes for a single amino acid, resulting in synthesis of polypeptide wherein the amino acid sequence is dictated by and matches the order of the codons in the mRNA. Sometimes, however, people speak of "translating" the DNA or RNA when they are merely reading the nucleotide sequence and predicting from it the sequence of the encoded protein. This might be more accurately termed "conceptual translation". Compare transcription.

Translational frame shifting: A mechanism used by certain viruses and even higher organisms to change the reading frame used during translation of an mRNA molecule, in a controlled manner, so that the polypeptide product is the result of translation in more than one reading frame.

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

Transposible elements: Genetic elements characterized by their abilities to insert into and withdraw from a given location within the genome, resulting in movement from site to site within the genome over a period of time. Transposable elements may cause epigenetic changes in phenotype.

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)

tRNA (transfer RNA): Special RNAs that are charged with amino acids and which carry anticodons for recognition of the codons in mRNA. tRNAs are responsible for translation from the language of nucleic acids into the language of polypeptides. Each tRNA can be charged with only one type of amino acid, although multiple tRNAs exist for many of the amino acids.

Tumor suppressor: A gene that prevents tumor formation until deleted or mutated. The best-known examples of tumor suppressors are the proteins p53 and Rb.

Turnover: The balance between synthesis and degradation of a product.

Universal Genetic Code: The "standard" codon usage that is common to most organisms.

Untranslated RNA: See Nontranslated RNA.

Upstream activator sequence: A binding site for transcription factors, generally part of a promoter region. A UAS may be found upstream of the TATA sequence (if there is one), and its function is (like an enhancer) to increase transcription. Unlike an enhancer, it can not be positioned just anywhere or in any orientation.

Upstream/Downstream: In an RNA, anything towards the 5' end of a reference point is "upstream" of that point. This orientation reflects the direction of both the synthesis of mRNA, and its translation - from the 5' end to the 3' end. In DNA, the situation is a bit more complicated. In the vicinity of a gene (or in a cDNA), the DNA has two strands, but one strand is virtually a duplicate of the RNA, so it's 5' and 3' ends determine upstream and downstream, respectively. NOTE that in genomic DNA, two adjacent genes may be on different strands and thus oriented in opposite directions. Upstream or downstream is only used on conjunction with a given gene.

Uracil: A nitrogenous base normally found in RNA but not DNA; uracil is capable of forming a base pair with adenine.

Vector: A construct used to propagate DNA in a host (bacteria, yeast, or cultured cells). The vector provides all sequences essential for replicating the test DNA. Typical vectors include plasmids, cosmids, phages and YACs.

Virion: A replication-competent virus particle.

Virulence: Ability to infect or cause disease.

Virus: A noncellular biological entity that can reproduce only within a host cell. Viruses consist of nucleic acid covered by protein; some animal viruses are also surrounded by membrane. Inside the infected cell, the virus uses the synthetic capability of the host to produce progeny virus.

VLSI: Very large- scale integration allowing over 100,000 transistors on a chip.

Western blot: A technique for analyzing mixtures of proteins to show the presence, size and abundance of one particular type of protein. Similar to Southern or Northern blotting (see "Blotting"), except that (1) a protein mixture is electrophoresed in an acrylamide gel, and (2) the "probe" is an antibody which recognizes the protein of interest, followed by a radioactive secondary probe (such as 125I-protein A).

Wildtype: The native or predominant genetic constitution before mutations, usually referring to the genetic consitution 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).

Xeno-immunization: Stimulation of immune response to antigens from different species.

YAC: See yeast artificial chromosome.

Yeast artificial chromosome (YAC): This is a method for cloning very large fragments of DNA. Genomic DNA in fragments of 200-500 kb are linked to sequences which allow them to propagate in yeast as a mini-chromosome (including telomeres, a centromere and an ARS - an autonomous replication sequence). This technique is used to clone large genes and intergenic regions, and for chromosome walking. Compare cloning vector, cosmid.

Z-DNA: Alternative left-handed form of the double helix which retains Watson-Crick type base pairing.

Zinc finger: A protein structural motif common in DNA binding proteins. Four Cys residues are found for each "finger" and one finger can bind a molecule of zinc. A typical configuration is: CysXxxXxxCys--(intervening 12 or so aa's)--CysXxxXxxCys.