Genetic engineering

Restriction enzymes (endonucleases)

BglII	5'…A_GATCT…3' (5'overhang)
BamHI	5'…G_GATCC…3' (5'overhang); notice: MboI and Sau3AI = 5'…_GATC…3'
EcoRI	5'…G_AATTC…3' (5' overhang)
HpaII	5'…C_CGG…3' (5' overhang); notice: CpG methylation blocked
HindIII	5'…A_AGCTT…3' (5' overhang)
KpnI	5'…GGTAC_C…3' (3' overhang)
NcoI	5'…C_CATGG…3' (5' overhang); notice: Kozak sequence (RCC)ACCATGG
NotI	5'…GC_GGCCGC…3' (5' overhang); notice: CpG methylation blocked
PstI	5'…CTGCA_G…3' (3' overhang)
SacI	5'…GAGCT_C…3' (3' overhang)
SalI	5'…G_TCGAC…3' (3' overhang); notice: compatible with XhoI
XhoI	5'…C_TCGAG…3' (5' overhang); notice: CpG me impaired, compatible with SalI
XmnI	5'…GAANN_NNTTC…3' (blunt end)

Restriction enzymes Restriction endonucleases are classified biochemically into three types, designated Type I, Type II and Type III. In type I and III systems, both the methylase and restriction activities are carried out by a single large enzyme complex. Although these enzymes recognize specific DNA sequences, the sites of actual cleavage are at variable distances from these recognition sites, and can be hundreds of bases away. In type II systems, the restriction enzyme is independent of its methylase, and cleavage occurs at very specific sites that are within or close to the recognition sequence. Type II enzymes are further classified according to their recognition site. Most type II enzymes cut palindromic DNA sequences, while type IIs recognise non-palindromic sequences and cleavage outside of the recognition site, and type IIb ones cut sequences twice at both sites outsite the recognition sequence.

Hosts for cloning

DNA cloning When cloning genomic DNA from organisms containing methylated bases, a strain should be used which lacks the appropriate methyl-specific restriction system(s). All mammals and higher plants, and many prokaryotes, contain methylcytosine. Bacteria and lower eukaryotes may contain methyladenine. However, Drosophilia melanogaster and Saccharomyces cerevisiae contain no methylated bases.

recA$^-$ host If the inserted DNA contains dispersed repeated sequences, recombination can occur between these, causing loss of pieces of the DNA. For plasmid libraries, this problem can be solved by propagating the DNA in a recA$^-$ host, where homologous recombination does not occur.

Vectors

Choosing a vector Size of the vector and the insert, copy number, polylinker (MCS), ability to select / screen for inserts.

Lambda FIX II Lambda replacement vector; allows inserts of 9-23kb; predigested with XhoI (5'C_TCGAG3'); DNA inserts should be digested with BamHI (5'G_GATCC3') or alternatively with MboI or Sau3AI (5'_GATC3'); fill-in the first two nucleotides of the restriction site (Klenow fragment + dATP + dGTP) in order to make it impossible for the insert to religate to itself. The stuffer fragment contains the red and gam genes for negative selection of vectors that did not take up an insert. Selection on host strains containing P2 phage lysogens. The polylinker of the lambda FIX II vector permits the isolation of the insert and flanking T3 and T7 bacteriophage promoters as an intact cassette by digestion with NotI.

pMAL-p2X The vector pMAL-p2X is designed to produce maltose-binding protein (MBP) fusions, where the protein of interest can be cleaved from MBP with the specific protease factor Xa.
A gene or ORF is inserted into a restriction site of the vector polylinker, in the same translational reading frame as the *mal*E gene (encoding maltose-binding protein). Insertion of the DNA fragment interrupts the *mal*E-*lac*Zα fusion pre-existing on the vector, affording a screen for inserts on the proper indicator plates. The fusion protein thus produced can be purified by amylose affinity chromatography.

Phagemid Plasmids that contain a filamentous phage origin of replication in addition to a plasmid ori. These phagemid vectors can be grown an propagated as plasmids. However, upon super-infection of a plasmid-containing cell with a wt helper phage, the phage ori becomes active and ssDNA is produced and secreted. There are usually (+) and (-) versions of these vectors where the phage ori is in opposite orientation so that it is possible to produce ssDNA from either DNA strand. The pBluescriptI, pBluescriptII, and pBS phagemid vectors derived from the general cloning vector pUC19 are examples of phagemids that incorporate the f1 filamentous phage ori.

Cosmid Cosmid vectors, plasmids containing a lambda phage cos site (e.g. pWE15), were developed to facilitate cloning of large DNA fragments. Commonly with ColE1 replicator, high copy number in E. coli. Cosmids can be packaged into lambda phage heads if the cos sites are separated by 40 to 50kb, the approximate length of the wt lambda genome.

BAC Plasmid cloning vector called bacterial artificial chromosome, developed using the F factor replicator for propagation of very large pieces of DNA (100-500kb).

lambda derived vectors About the middle third of the lambda genome is dispensable for lytic growth. Derivatives of phage lambda that are used as cloning vectors typically contain restriction enzyme sites flanking some or all of these dispensable genes. The major advantage of using lambda-derived cloning vectors is that DNA can be inserted and packaged into phages in vitro. Although the efficiency of packaging only approaches 10%, phages, once package, form plaques on E. coli with an efficiency of 1. Packaging occurs only if the length of the recombinant lambda phage is > 78% and < 105% of wt lambda DNA.

Special sequences

Shine – Dalgarno Only in prokaryotes mRNAs; 6nt, AGGAGG; 4-7nt upstream of the initiation AUG codon; base pairs with the 3’ end of 16S rRNA (small ribosomal subunit).

Kozak sequence Only in eukaryotes; (RCC)**A**CC*ATG***G**, R = purine; short recognition sequence that greatly facilitates initial binding of mRNA to the small subunit of the ribosome (5’ cap helps as well in eukaryotes!). NcoI cleave this site (5’C_CATGG3’).

STS Sequence tagged site. Short (200 to 500bp) DNA sequence that has a single occurrence in the genome and whose location and base sequence are known. Detectable by PCR, STSs are useful for localizing and orienting the mapping and sequence data and serve as landmarks on physical maps of genomes.

EST Expressed sequence tag. A unique segment of cDNA with a base sequence identical to at least part of the coding region of a gene, generally used as landmark for mapping. ESTs are STSs derived from cDNAs.

Gels

Agarose gel Revelation with EtBr or other dye. Usually used to detect high / mid molecular weight DNA / RNA, although at 4% of agarose DNA smaller than 30bp can be seen. Normally, gels are run at 1,5 - 3%. Agarose gels are easier to handle, cheaper and the separation is quicker.

Acrylamide gel Revelation with Coomassie blue (for proteins) or $AgNO_3$. Used for both protein and nucleic acid electrophoresis. Higher resolution than agarose gels, single bp differences can be detected. Normally run at 8 - 12%.

Techniques

TdT Terminal deoxynucleotidyl transferase is a template independent polymerase that catalyses the addition of deoxynucleotides to the 3' hydroxyl terminus of DNA molecules. Protruding, recessed or blunt-ended double or single-stranded DNA molecules serve as a substrate for TdT.
Applications:
- Addition of homopolymer tails to the 3' ends of DNA.
- Labelling the 3' ends of DNA.

Homopolymer tailing Attachment of a sequence of identical nucleotides (e.g. TTTTT) to the end of a nucleic acid molecule. Typically, single-stranded homopolymer extensions are added by the TdT to the ends of a double-stranded DNA molecule.

S1 nuclease A single-strand specific nuclease, degrades single-stranded nucleic acids, releasing 5'-phosphoryl mono- or oligonucleotides. Five times more active on DNA than on RNA.
Applications:
- Removal of single-stranded overhangs of DNA fragments.
- RNA transcript mapping.
- Creation of unidirectional deletions in DNA fragments in conjunction with exonuclease III.

RNAseH Enzyme that specifically snips the RNA in DNA-RNA hybrids. The produced RNA fragments can serve as primers for DNA polymerase.

Alkaline phosphatase Removes 5' phosphate groups from DNA and RNA. It will also remove phosphates from nucleotides and proteins. These enzymes are most active at alkaline pH - hence the name.
Applications:
- Removing 5' phosphates from plasmid and bacteriophage vectors that have been cut with a restriction enzyme. In subsequent ligation reactions, this treatment prevents self-ligation of the vector and thereby greatly facilitates ligation of other DNA fragments into the vector (e.g. subcloning).
- Removing 5' phosphates from fragments of DNA prior to labeling with radioactive phosphate. Polynucleotide kinase is much more effective in phosphorylating DNA if the 5' phosphate has previously been removed.

Klenow fragment A proteolytic fragment of DNA polymerase I of E. coli that contains the DNA polymerase activity and the 3' to 5' proofreading activity but lacks the 5' to 3' exonuclease activity. This enzyme is widely used for DNA sequencing.
Applications:
- Synthesis of double-stranded DNA from single-stranded templates. DNA polymerases require a primer to provide a free 3' hydroxyl group for initiation of synthesis and dNTPs.
- Filling in recessed 3' ends of DNA fragments. A "fill-in" reaction is used to create blunt ends on fragments created by cleavage with restriction enzymes that leave 5' overhangs. Conceptually identical to the reaction above.
- Digesting away protruding 3' overhangs. The 3' -> 5' exonuclease activity of Klenow will digest away the protruding overhang. Removal of nucleotides from the 3' ends will continue, but, in the presence of nucleotides, the polymerase activity will balance the exonuclease activity, yielding blunt ends. This reaction is more efficienty conducted with T4 DNA polymerase, which has much more potent exonuclease activity.
- Preparation of radioactive DNA probes. Use radioactive dNTPs.

T4 DNA polymerase T4 is a bacteriophage of E. coli. The activities of T4 DNA polymerase are very similar to those of the Klenow fragment - it functions as a 5' -> 3' DNA polymerase and a 3' -> 5' exonuclease, but does not have 5' -> 3' exonuclease activity. In general, T4 DNA polymerase is used for the same types of reactions as Klenow fragment.
The two major differences are:

- The 3' -> 5' exonuclease activity of T4 DNA polymerase is roughly 200 times that of Klenow fragment, making it preferred by many investigators for blunting DNAs with 3' overhangs.
- While Klenow fragment will displace downstream oligonucleotides as it polymerizes, T4 DNA polymerase will not.

CAT enzyme assay Chloramphenicol Acetyltransferase (CAT) comes from microorganisms and inactivates chloramphenicol by catalysing the transfer of acetyl groups from acetyl coenzyme A to chloramphenicol. With the CAT assay, the CAT-containing lysates of transfected cells are incubated with ^{14}C-chloramphenicol, which is then acetylated. Acetylated and non-acetylated ^{14}C-chloramphenicol is then separated using thin-layer chromatography and visualized by autoradiography. This gene is not found in eukaryotes, and therefore eukaryotic cells contain no background of CAT activity. As an increasingly common, non-radioactive alternative, CAT expression is quantified by an ELISA via immunological detection of the CAT enzyme which has been expressed.
Application:

- The pCAT®3 Reporter Vectors (Promega) provide a basis for the quantitative analysis of factors that may regulate mammalian gene expression.

ß-Galactosidase The prokaryotic ß-galactosidase naturally catalyzes the hydrolysis of ß-galactosides (e.g. lactose). However, the use of non-physiological substrates also enables the quantification of ß-galactosidase activity in lysates of transfected cells via spectrophotometry (e.g. with 0-nitrophenyl-ß-D-galactoside = ONPG), fluorometry (e.g. with a 4-methyl-umbelliferyl-ß-galactopyranoside compound = MUG) or via chemiluminescence. Detection by chemiluminescence (e.g. with 1,2 dioxetan-galactopyranoside derivatives) is 100–1000 times more sensitive than the other two detection methods, and thus even more sensitive than the luciferase assay. A major advantage of this system is the fact that ß-galactosidase activity can also be measured in situ.

GFP Unlike other bioluminescent reporters, the green fluorescent protein from the Aequorea victoria jellyfish requires no additional proteins, substrates or co-factors to emit light. When irradiated with UV light or blue light, it emits green light, which enables the examination of gene expression and protein localization in situ and in vivo. In addition, the gene expression can be observed in real time. However, the system is less suitable for quantifying the gene expression.

SAGE Serial analysis of gene expression (SAGE) is a powerful tool that allows the analysis of overall gene expression patterns with digital analysis. Because SAGE does not require a preexisting clone, it can be used to identify and quantitate new genes as well as known genes.
The basic concept of SAGE rests on two principles: firstly, a small sequence of nucleotides from the transcript, called a 'tag', can effectively identify the original transcript from whence it came, and secondly, that linking these tags allows for rapid sequencing analysis of multiple transcripts

Figure 1 shows a schematic diagram of each of the steps in SAGE.

- First, a cDNA of each transcript in the cell must be generated. This is necessary, since mRNA is much less stable than DNA. A biotinylated oligo dT primer is used that will complimentary base pair with the poly(A) tail of every mRNAs in a cell. RT then synthesises a DNA strand that is complimentary to the mRNA.
- This DNA strand will then be converted to a double-stranded DNA molecule, which can then proceed to the next step.
- Once the cDNA has been created, it is then cleaved using an anchoring enzyme. The anchoring enzyme is a restriction endonuclease that recognizes and cuts specific 4bp DNA sequences (Average cleavage every 256bp). The cut cDNA is then immobilized on streptavidin beads.
- The sample of bound cDNAs is then divided in half and ligated to either linker A or B. These linkers are designed to contain a type IIS restriction site. Type IIS restriction endonucleases cut at a defined distance up to 20 base pairs away from their recognition sites. The type IIS restriction endonuclease,

also called the tagging enzyme, cleaves the cDNA to release it from its bound bead. Blunt ends are then created.

- Once this is achieved, the cDNA tags bound to linker A and B are ligated to each other to create ditags. These ditags are then amplified by PCR, using primers that are complimentary to sequence in either linker.
- Once the ditags have been amplified, they are then cleaved using the anchoring enzyme again. This releases the linkers from either end of the ditag and creates sticky ends. In this way, all of the ditags generated are concatenated to produce one long string of tags.
- This collection of tags is then introduced into a vector to be cloned and sequenced.
- The tags are then quantified and the patterns of gene expression determined (comparison between normal and disease for example).

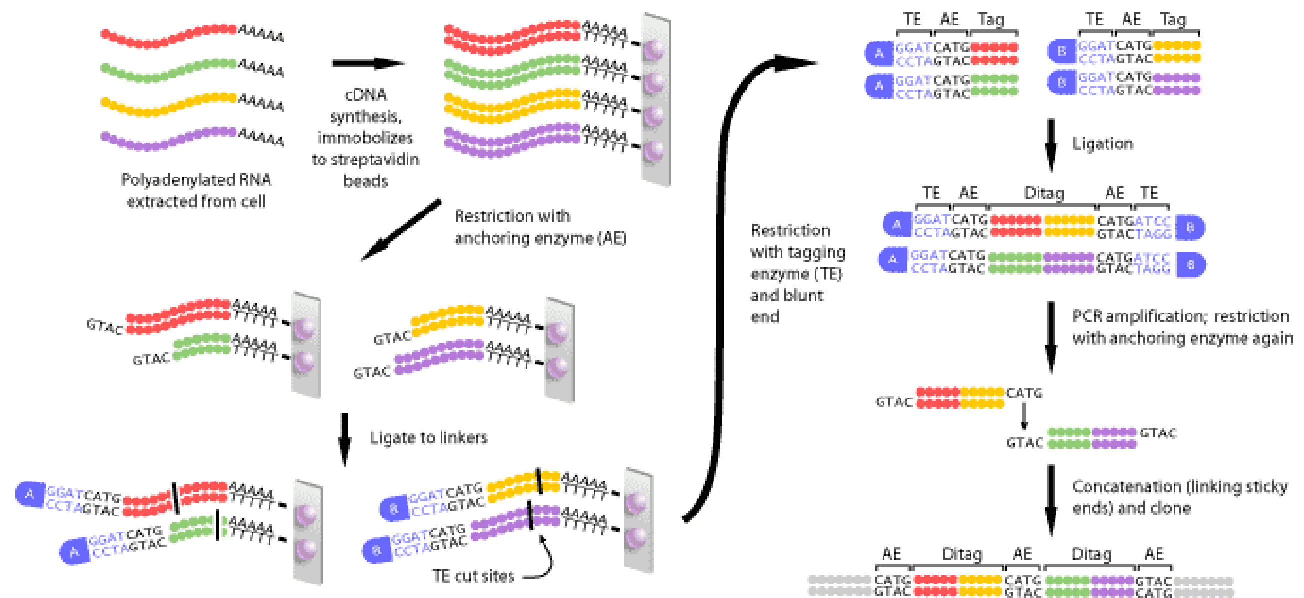

Figure 1. The steps involved in SAGE.

Methods and protocols

DNA preparation Purification and concentration of genomic DNA from small volumes (< 0,4ml) at concentrations < 1mg/ml. The DNA solution is first extracted with a phenol / chloroform / isoamyl alcohol mixture to remove protein contaminants, then precipitated with 100% ethanol. The DNA is pelleted after the precipitation step, washed with 70% ethanol to remove salts and small organic molecules, and resuspended in buffer at a concentration suitable for further experimentation.
Alternatively, anion exchange chromatography techniques can be applied to purify DNA. A crude nucleic acid sample (usually a cell lysate) is applied to the (Qiagen) column under conditions that favour binding. Contaminants in the sample are washed from the column with a moderate-salt buffer and DNA eluted using a high-salt buffer.
Another possibility is the silica-gel membrane technology (Qiagen). Different kits are available for various DNA source types (e.g. DNeasy Tissue Kit for animal tissues and cells, yeast, and bacteria). Principle: The DNeasy Tissue Kit uses advanced silica-gel membrane technology for rapid and efficient isolation of total cellular DNA without phenol, chloroform, or ethanol precipitation. Mechanical homogenisation is not necessary since tissues are directly lysed by proteinase K. The buffer system is optimised to allow selective binding of DNA to the DNeasy membrane. Simple centrifugation processing completely removes contaminants and enzyme inhibitors such as proteins, divalent cations, and secondary metabolites. Procedure: Lysates are loaded onto DNeasy spin columns. During a brief spin the DNA selectively binds to the silica-gel membrane while contaminants pass through. Remaining contaminants and enzyme inhibitors are removed in two efficient wash steps. Pure DNA is then eluted in water or low-salt buffer, ready for use. Time after lysis: 20min.

cDNA synthesis Use the ZAP Express cDNA Synthesis Kit (Stratagene) based on vectors or the StrataScript QPCR cDNA Synthesis Kit designed for conversion of RNA to cDNA and fully optimized for two-step quantitative reverse transcription-PCR (QRT-PCR) applications. A two-step RT-PCR format is useful for amplifying multiple targets from a single cDNA source, for maintaining archival cDNA, and for providing maximum flexibility in selecting a downstream QPCR reagent system. 15-minute cDNA synthesis step and an overall protocol time of just 25 minutes. The kit is formatted for high-yield production of cDNA up to 12 kb and each reaction accommodates a range of RNA amounts from fg to µg.
cDNA templates can be generated from either poly(A) mRNA or total RNA. Key to the performance of the kit is the QPCR-grade StrataScript Reverse Transcriptase (RT) enzyme and the corresponding QPCR-optimized RT buffer. The cDNA priming strategy can affect cDNA yield, sensitivity, and detection of certain targets, such as GC-rich targets or sequences located at the 5' or 3' end of a transcript. For this reason, individually packaged random nonamers and oligo-dT primers are provided separately from the master mix, allowing to use the best priming strategy for a specific target. The Brilliant QPCR master mix is designed for use with molecular beacons or TaqMan probes.

DNA enrichment Genomic DNA library enrichment allows to amplify specific DNA fragments of a library before ligating them in vectors for cloning and sequencing. The technique is based on streptavidin-coated paramagnetic beads that are used to isolate biotin-labelled probes that bind to target molecules. The stability and strength of the streptavidin-biotin interaction enable DNA manipulations such as strand melting, hybridisation and elution to be performed without affecting the immobilization of the DNA on the coated magnetic beads. Protocol:

- Genomic DNA of good quality is digested by appropriate restriction endonucleases to give DNA fragments of suitable size.
- Ligation of adaptors
- Preparation of a specific probe for the DNA fragment to enrich
- Isolation of the target molecule by a magnet (magnet-paramagnetic beads-streptavidin-biotin-probe-target DNA fragment)
- Library amplification of the target molecule with PCR (primers on adaptors) and subsequent control on an agarose gel with revelation by EtBr.
- If cloning is desired, preparation of a suitable plasmid (RE sites, size) and cloning into this. Then transformation or infection of host cells.

SSH Suppression subtractive hybridisation is a highly effective method that has been developed to detect rare differential gene expression between a tester population (with ds cDNA of interest) and a driver population. This method allows as well the generation of differentially regulated or tissue-specific cDNA probes and libraries. It is primarily based on suppression PCR and combines normalisation and subtraction in a single procedure. The normalisation step equalizes the abundance of cDNAs within the target population and the subtraction step excludes the common sequences between the target and driver populations (both ds cDNA populations).