Polymerase chain reaction

From Wikipedia, the free encyclopedia

• *Ten things you may not know about images on Wikipedia* •

"PCR" redirects here. For other uses, see PCR (disambiguation).

A strip of eight PCR tubes, each tube contains a 100µl reaction.

The **polymerase chain reaction** (**PCR**) is a technique widely used in molecular biology. It derives its name from one of its key components, a DNA polymerase used to amplify (i.e., replicate) a piece of DNA by *in vitro* enzymatic replication. As PCR progresses, the DNA thus generated is itself used as template for replication. This sets in motion a chain reaction in which the DNA template is exponentially amplified. With PCR it is possible to amplify a single or few copies of a piece of DNA across several orders of magnitude, generating millions or more copies of the DNA piece. PCR can be performed without restrictions on the form of DNA, and it can be extensively modified to perform a wide array of genetic manipulations.

Almost all PCR applications employ a heat-stable DNA polymerase, such as Taq polymerase. This DNA polymerase enzymatically assembles a new DNA strand from DNA building blocks, the nucleotides, using single-stranded DNA as template and DNA oligonucleotides (also called DNA primers) required for initiation of DNA synthesis. The vast majority of PCR methods use thermal cycling, i.e., alternately heating and cooling the PCR sample to a defined series of temperature steps. These different temperature steps are necessary to bring about physical separation of the strands in a DNA double helix (DNA melting), and permit DNA synthesis by the DNA polymerase to selectively amplify the target DNA. The power and selectivity of PCR are primarily due to selecting primers that are highly complementary to the DNA region targeted for amplification, and to the thermal cycling conditions used.

Developed in 1983 by Kary Mullis, PCR is now a common and often indispensable technique used in medical and biological research labs for a variety of applications. These include DNA cloning for sequencing, DNA-based phylogeny, or functional analysis of genes; the diagnosis of hereditary diseases; the identification of genetic fingerprints (used in forensics and paternity testing); and the detection and diagnosis of infectious diseases. Mullis won the Nobel Prize for his work on PCR.[1]

Contents

- 1 PCR principle and procedure
  - 1.1 Procedure
  - 1.2 PCR optimization
- 2 Application of PCR
  - 2.1 Isolation of genomic DNA
  - 2.2 Amplification and quantitation of DNA
- 3 Variations on the basic PCR technique
- 4 History
  - 4.1 Patent wars
- 5 References
- 6 External links

PCR principle and procedure

Figure 1a: An old thermal cycler for PCR

Figure 1b: A very old three-temperature thermal cycler for PCR

PCR is used to amplify specific regions of a DNA strand (the DNA target). This can be a single gene, a part of a gene, or a non-coding sequence. Most PCR methods typically amplify DNA fragments of up to 10 kilo base pairs (kb), although some techniques allow for amplification of fragments up to 40 kb in size.[2]

A basic PCR set up requires several components and reagents. [3] These components include:

- *DNA template* that contains the DNA region (target) to be amplified.
- One or more *primers,* which are complementary to the DNA regions at the 5' (five prime) and 3' (three prime) ends of the DNA region.
- a DNA polymerase such as *Taq polymerase* or another DNA polymerase with a temperature optimum at around 70°C.
- *Deoxynucleotide triphosphates* (dNTPs), the building blocks from which the DNA polymerases synthesizes a new DNA strand.
- *Buffer solution,* providing a suitable chemical environment for optimum activity and stability of the DNA polymerase.
- *Divalent cations,* magnesium or manganese ions; generally Mg^{2+} is used, but Mn^{2+} can be utilized for PCR-mediated DNA mutagenesis, as higher Mn^{2+} concentration increases the error rate during DNA synthesis [4]
- *Monovalent cation* potassium ions.

The PCR is commonly carried out in a reaction volume of 15-100 µl in small reaction tubes (0.2-0.5 ml volumes) in a thermal cycler. The thermal cycler allows heating and cooling of the reaction tubes to control the temperature required at each reaction step (see below). Thin-walled reaction tubes permit favorable thermal conductivity to allow for rapid thermal equilibration. Most thermal cyclers have heated lids to prevent condensation at the top of the reaction tube. Older thermocyclers lacking a heated lid require a layer of oil on top of the reaction mixture or a ball of wax inside the tube.

Procedure

Figure 2: Schematic drawing of the PCR cycle. **(1) Denaturing at 94-96°C. (2) Annealing at ~65°C (3) Elongation at 72°C**. Four cycles are shown here.

The PCR usually consists of a series of 20 to 35 repeated temperature changes called cycles; each cycle typically consists of 2-3 discrete temperature steps. Most commonly PCR is carried out with cycles that have three temperature steps (Fig. 2). The cycling is often preceded by a single temperature step (called *hold*) at a high temperature (>90°C), and followed by one hold at the end for final product extension or brief storage. The temperatures used and the length of time they are applied in each cycle depend on a variety of parameters. These include the enzyme used for DNA synthesis, the concentration of divalent ions and dNTPs in the reaction, and the melting temperature (Tm) of the primers. [5]

- *Initialization step*: This step consists of heating the reaction to a temperature of 94-96°C (or 98°C if extremely thermostable polymerases are used), which is held for 1-9 minutes. It is only required for DNA polymerases that require heat activation by hot-start PCR. [6]

- *Denaturation step*: This step is the first regular cycling event and consists of heating the reaction to 94-98°C for 20-30 seconds. It causes melting of DNA template and primers by disrupting the hydrogen bonds between complementary bases of the DNA strands, yielding single strands of DNA.

- *Annealing step*: The reaction temperature is lowered to 50-65°C for 20-40 seconds allowing annealing of the primers to the single-stranded DNA template. Typically the annealing temperature is about 3-5 degrees Celsius below the Tm of the primers used. Stable DNA-DNA hydrogen bonds are only formed when the primer sequence very closely matches the template sequence. The polymerase binds to the primer-template hybrid and begins DNA synthesis.

- *Extension/elongation step*: The temperature at this step depends on the DNA polymerase used; Taq polymerase has its optimum activity temperature at 75-80°C, [7][8] and commonly a temperature of 72°C is used with this enzyme. At this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTP's that are complementary to the template in 5' to 3' direction, condensing the 5'-phosphate group of the dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. The extension time depends both on the DNA polymerase used and on the length of the DNA fragment to be amplified. As a rule-of-thumb, at its optimum temperature, the DNA polymerase will polymerize a thousand bases in one minute.

- *Final elongation*: This single step is occasionally performed at a temperature of 70-74°C for 5-15 minutes after the last PCR cycle to ensure that any remaining single-stranded DNA is fully extended.

- *Final hold*: This step at 4-15°C for an indefinite time may be employed for short-term storage of the reaction.

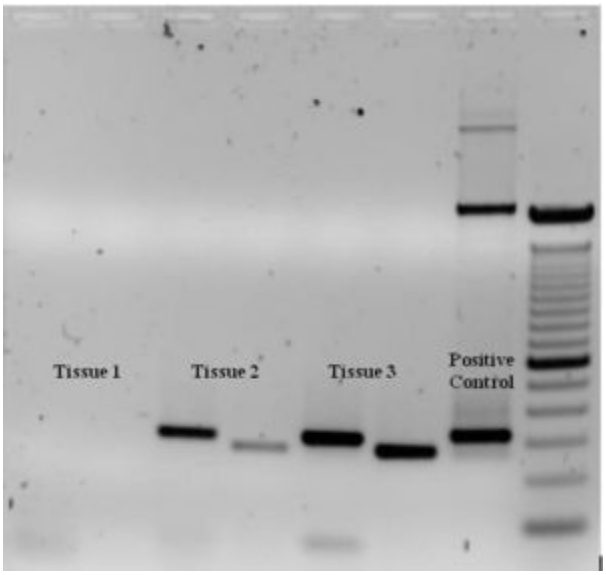


Figure 3: Ethidium bromide-stained PCR products after gel electrophoresis. Two sets of primers were used to amplify a target sequence from three different tissue samples. No amplification is present in sample #1; DNA bands in sample #2 and #3 indicate successful amplification of the target sequence. The gel also shows a positive control, and a DNA ladder containing DNA fragments of defined length for sizing the bands in the experimental PCRs.

To check whether the PCR generated the anticipated DNA fragment (also sometimes referred to as amplimer), agarose gel electrophoresis is employed for size separation of the PCR products. The size(s) of PCR products is determined by comparison with a *DNA ladder*, which contains DNA fragments of known size, run on the gel alongside the PCR products (see Fig. 3).

PCR optimization

Main article: PCR optimization

In practice, PCR can fail for various reasons, in part due to its sensitivity to contamination causing amplification of spurious DNA products. Because of this, a number of techniques and procedures have been developed for optimizing PCR conditions. Contamination with extraneous DNA is addressed with lab protocols and procedures that separate pre-PCR reactions from potential DNA contaminants.[3] This usually involves spatial separation of PCR-setup areas from areas for analysis or purification of PCR products, and thoroughly cleaning the work surface between reaction setups. Primer-design techniques are important in improving PCR product yield and in avoiding the formation of spurious products, and the usage of alternate buffer components or polymerase enzymes can help with amplification of long or otherwise problematic regions of DNA.

Application of PCR

Isolation of genomic DNA

PCR allows isolation of DNA fragments from genomic DNA by selective amplification of a specific region of DNA. This use of PCR augments many methods, such as Southern and northern blotting and DNA cloning, that require large amounts of DNA, representing a specific DNA region. PCR supplies these techniques with high amounts of pure DNA, enabling analysis of DNA samples even from very small amounts of starting material.

Other applications of PCR include DNA sequencing to determine unknown PCR-amplified sequences in which one of the amplification primers may be used in Sanger sequencing, isolation of a DNA sequence to expedite recombinant DNA technologies involving the insertion of a DNA sequence into a plasmid or the genetic material of another organism. PCR may also be used for genetic fingerprinting a forensic technique used to identify a person or organism by comparing experimental DNAs. There are different PCR-based methods for fingerprinting, summarized in *Genetic fingerprinting*.

PCR 'fingerprints' are unique and genetic relationships, for example, parent-child or siblings, can be determined from two or more genetic fingerprints, which can thus be used for paternity tests (Fig. 4). A variation of this technique can also be used to determine evolutionary relationships among organisms.

Figure 4: Electrophoresis of PCR-amplified DNA fragments. (1) Father. (2) Child. (3) Mother. The child has inherited some, but not all of the fingerprint of each of its parents, giving it a new, unique fingerprint.

Amplification and quantitation of DNA

Because PCR amplifies the regions of DNA that it targets, PCR can be used to analyze extremely small amounts of sample. This is often critical for forensic analysis, when only a trace amount of DNA is available as evidence. PCR may also be used in the analysis of ancient DNA that is thousands of years old. These PCR-based techniques have been successfully used on animals, such as a forty-thousand-year-old mammoth, and also on human DNA, in applications ranging from the analysis of Egyptian mummies to the identification of a Russian Tsar[9].

Viral DNA can likewise be detected by PCR. The primers used need to be specific to the targeted sequences in the DNA of a virus, and the PCR can be used for diagnostic analyses or DNA sequencing of the viral genome. The high sensitivity of PCR permits virus detection soon after infection and even before the onset of disease. Such early detection may give physicians a significant lead in treatment. The amount of virus ("viral load") in a patient can also be quantified by PCR-based DNA quantitation techniques (see below).

Quantitative PCR methods allow the estimation of the amount of a given sequence present in a sample – a technique often applied to quantitatively determine levels of gene expression. Real-time PCR is an established tool for DNA quantification that measures the accumulation of DNA product after each round of PCR amplification.

Variations on the basic PCR technique

- **Allele-specific PCR**: This diagnostic or cloning technique is used to identify or utilize single-nucleotide polymorphisms (SNPs) (single base differences in DNA). It requires prior knowledge of a DNA sequence, including differences between alleles, and uses primers whose 3' ends encompass the SNP. PCR amplification under stringent conditions is much less efficient in the presence of a mismatch between template and primer, so successful amplification with a SNP-specific primer signals presence of the specific SNP in a sequence. [10]

- **Assembly PCR**: Assembly PCR is the artificial synthesis of long DNA sequences by performing PCR on a pool of long oligonucleotides with short overlapping segments. The oligonucleotides alternate between sense and antisense directions, and the overlapping segments determine the order of the PCR fragments thereby selectively producing the final long DNA product. [11]

- **Asymmetric PCR**: Asymmetric PCR is used to preferentially amplify one strand of the original DNA more than the other. It finds use in some types of sequencing and hybridization probing where having only one of the two complementary stands is required. PCR is carried out as usual, but with a great excess of the primers for the chosen strand. Due to the slow (arithmetic) amplification later in the reaction after the limiting primer has been used up, extra cycles of PCR are required. [12] A recent modification on this process, known as **L**inear-**A**fter-**T**he-**E**xponential-PCR (LATE-PCR), uses a limiting primer with a higher melting temperature (Tm) than the excess primer to maintain reaction efficiency as the limiting primer concentration decreases mid-reaction. [13]

- **Colony PCR**: Bacterial colonies (E.coli) can be rapidly screened by PCR for correct DNA vector constructs. Selected bacterial colonies are picked with a sterile toothpick and dabbed into the PCR master mix or sterile water. The PCR is started with an extended time at 95˚C when standard polymerase is used or with a shortened denaturation step at 100˚C and special chimeric DNA polymerase.[14]

- **Helicase-dependent amplification**: This technique is similar to traditional PCR, but uses a constant temperature rather than cycling through denaturation and annealing/extension cycles. DNA Helicase, an enzyme that unwinds DNA, is used in place of thermal denaturation.[15]

- **Hot-start PCR**: This is a technique that reduces non-specific amplification during the initial set up stages of the PCR. The technique may be performed manually by heating the reaction components to the melting temperature (e.g., 95˚C) before adding the polymerase.[16] Specialized enzyme systems have been developed that inhibit the polymerase's activity at ambient temperature, either by the binding of an antibody[6] or by the presence of covalently bound inhibitors that only dissociate after a high-temperature activation step. Hot-start/cold-finish PCR is achieved with new hybrid polymerases that are inactive at ambient temperature and are instantly activated at elongation temperature.

- **Intersequence-specific (ISSR) PCR**: a PCR method for DNA fingerprinting that amplifies regions between some simple sequence repeats to produce a unique fingerprint of amplified fragment lengths.[17]

- **Inverse PCR**: a method used to allow PCR when only one internal sequence is known. This is especially useful in identifying flanking sequences to various genomic inserts. This involves a series of DNA digestions and self ligation, resulting in known sequences at either end of the unknown sequence. [18]

- **Ligation-mediated PCR**: This method uses small DNA linkers ligated to the DNA of interest and multiple primers annealing to the DNA linkers; it has been used for DNA sequencing, genome walking, and DNA footprinting. [19]

- **Methylation-specific PCR** (MSP): The MSP method was developed by Stephen Baylin and Jim Herman at the Johns Hopkins School of Medicine, [20] and is used to detect methylation of CpG islands in genomic DNA. DNA is first treated with sodium bisulfite, which converts unmethylated cytosine bases to uracil, which is recognized by PCR primers as thymine. Two PCR reactions are then carried out on the modified DNA, using primer sets identical except at any CpG islands within the primer sequences. At these points, one primer set recognizes DNA with cytosines to amplify methylated DNA, and one set recognizes DNA with uracil or thymine to amplify unmethylated DNA. MSP using qPCR can also be performed to obtain quantitative rather than qualitative information about methylation.

- **Multiplex Ligation-dependent Probe Amplification** (**MLPA**): permits multiple targets to be amplified with only a single primer pair, thus avoiding the resolution limitations of multiplex PCR (see below).

- **Multiplex-PCR**: The use of multiple, unique primer sets within a single PCR reaction to produce amplicons of varying sizes specific to different DNA sequences. By targeting multiple genes at once, additional information may be gained from a single test run that otherwise would require several times the reagents and more time to perform. Annealing temperatures for each of the primer sets must be optimized to work correctly within a single reaction, and amplicon sizes, i.e., their base pair length, should be different enough to form distinct bands when visualized by gel electrophoresis.

- **Nested PCR**: increases the specificity of DNA amplification, by reducing background due to non-specific amplification of DNA. Two sets of primers are being used in two successive PCR reactions. In the first reaction, one pair of primers is used to generate DNA products, which besides the intended target, may still consist of non-specifically amplified DNA fragments. The product(s) are then used in a second PCR reaction with a set of primers whose binding sites are completely or partially different from and located 3' of each of the primers used in the first reaction. Nested PCR is often more successful in specifically amplifying long DNA fragments than conventional PCR, but it requires more detailed knowledge of the target sequences.

- **Quantitative PCR (Q-PCR)**: is used to measure the quantity of a PCR product (preferably real-time). It is the method of choice to quantitatively measure starting amounts of DNA, cDNA or RNA. Q-PCR is commonly used to determine whether a DNA sequence is present in a sample and the number of its copies in the sample. The method with currently the highest level of accuracy is **Quantitative real-time PCR**. It is often confusingly known as RT-PCR (**R**eal **T**ime PCR) or RQ-PCR. QRT-PCR or RTQ-PCR are more appropriate contractions. RT-PCR commonly refers to reverse transcription PCR (see below), which is often used in conjunction with Q-PCR. QRT-PCR methods use fluorescent dyes, such as Sybr Green, or fluorophore-containing DNA probes, such as TaqMan, to measure the amount of amplified product in real time.

- **RT-PCR**: (**R**everse **T**ranscription PCR) is a method used to amplify, isolate or identify a known sequence from a cellular or tissue RNA. The PCR is preceded by a reaction using reverse transcriptase to convert RNA to cDNA. RT-PCR is widely used in expression profiling, to determine the expression of a gene or to identify the sequence of an RNA transcript, including transcription start and termination sites and, if the genomic DNA sequence of a gene is known, to map the location of exons and introns in the gene. The 5' end of a gene (corresponding to the transcription start site) is typically identified by a RT-PCR method, named RACE-PCR, short for *Rapid Amplification of cDNA Ends*.

- **TAIL-PCR**: **Thermal asymmetric interlaced** PCR is used to isolate unknown sequence flanking a known sequence. Within the known sequence TAIL-PCR uses a nested pair of primers with differing annealing temperatures; a degenerate primer is used to amplify in the other direction from the unknown sequence.[21]

- **Touchdown PCR**: a variant of PCR that aims to reduce nonspecific background by gradually lowering the annealing temperature as PCR cycling progresses. The annealing temperature at the initial cycles is usually a few degrees (3-5˚C) above the T_m of the primers used, while at the later cycles, it is a few degrees (3-5˚C) below the primer T_m. The higher temperatures give greater specificity for

primer binding, and the lower temperatures permit more efficient amplification from the specific products formed during the initial cycles. [22]

- **PAN-AC**: This method uses isothermal conditions for amplification, and may be used in living cells.[23][24].

History

A 1971 paper in the Journal of Molecular Biology by Kleppe and co-workers first described a method using an enzymatic assay to replicate a short DNA template with primers in vitro. [25] However, this early manifestation of the basic PCR principle did not receive much attention, and the invention of the polymerase chain reaction in 1983 is generally credited to Kary Mullis [26][27]. He was awarded the Nobel Prize in Chemistry in 1993 for his invention,[1] seven years after he and his colleagues at Cetus first put his proposal to practice. However, some controversies have remained about the intellectual and practical contributions of other scientists to Mullis' work, and whether he had been the sole inventor of the PCR principle. (see main article: Kary Mullis)
At the time he developed PCR in 1983, Mullis was working in Emeryville, California for Cetus Corporation, one of the first biotechnology companies. There, he was responsible for synthesizing short chains of DNA. Mullis has written that he conceived of PCR while cruising along the Pacific Coast Highway one night in his car[26]. He was playing in his mind with a new way of analyzing changes (mutations) in DNA when he realized that he had instead invented a method of amplifying any DNA region through repeated cycles of duplication driven by an enzyme called DNA polymerase. Mullis credits the psychedelic drug LSD for his invention of the technique. [1](*Video*)
In *Scientific American*, Mullis summarized the accomplishment: "Beginning with a single molecule of the genetic material DNA, the PCR can generate 100 billion similar molecules in an afternoon. The reaction is easy to execute. It requires no more than a test tube, a few simple reagents, and a source of heat."[28]

DNA polymerase occurs naturally in living organisms. In cells it functions to duplicate DNA when cells divide in mitosis and meiosis. Polymerase works by binding to a single DNA strand and creating the complementary strand. In the first of many original processes, the enzyme was used in vitro (in a controlled environment outside an organism). The double-stranded DNA was separated into two single strands by heating it to 94°C (201°F). At this temperature, however, the DNA polymerase used at the time were destroyed, so the enzyme had to be replenished after the heating stage of each cycle. The original procedure was very inefficient, since it required a great deal of time, large amounts of DNA polymerase, and continual attention throughout the process.

In 1986, this original PCR process was greatly improved by the use of DNA polymerase taken from thermophilic bacteria grown in geysers at a temperature of over 110°C (230°F). The DNA polymerase taken from these organisms is stable at high temperatures and, when used in PCR, does not break down when the mixture was heated to separate the DNA strands. Since there was no longer a need to add new DNA polymerase for each cycle, the process of copying a given DNA strand could be simplified and automated.

One of the first thermostable DNA polymerases was obtained from *Thermus aquaticus* and was called "Taq." Taq polymerase is widely used in current PCR practice. A disadvantage of Taq is that it sometimes makes mistakes when copying DNA, leading to mutations (errors) in the DNA sequence, since it lacks 3'→5' proofreading exonuclease activity. Polymerases such as *Pwo* or *Pfu,* obtained from *Archaea,* have *proofreading mechanisms* (mechanisms that check for errors) and can significantly reduce the number of mutations that occur in the copied DNA sequence. However these enzymes polymerise DNA at a much slower rate than Taq. Combinations of both *Taq* and *Pfu* are available nowadays that provide both high processivity (fast polymerisation) and high fidelity (accurate duplication of DNA).

PCR has been performed on DNA larger than 10 kilobases, but the average PCR is only several hundred to a few thousand bases of DNA. The problem with long PCR is that there is a balance between accuracy and processivity of the enzyme. Usually, the longer the fragment, the greater the probability of errors.

Patent wars

The PCR technique was patented by Cetus Corporation, where Mullis worked when he invented the technique in 1983. The *Taq* polymerase enzyme was also covered by patents. There have been several high-profile lawsuits related to the technique, including an unsuccessful lawsuit brought by DuPont. The pharmaceutical company Hoffmann-La Roche purchased the rights to the patents in 1992 and currently holds those that are still protected.

A related patent battle over the Taq polymerase enzyme is still ongoing in several jurisdictions around the world between Roche and Promega. Interestingly, it seems possible that the legal arguments will extend beyond the life of the original PCR and Taq polymerase patents, which expired on March 28, 2005.[29]

References

1. ^ [a] [b] Karry Mullis Nobel Lecture, December 8, 1993
2. ^ Cheng S, Fockler C, Barnes WM, Higuchi R (1994). "Effective amplification of long targets from cloned inserts and human genomic DNA". *Proc Natl Acad Sci.* **91**: 5695-5699. PMID 8202550.
3. ^ [a] [b] Joseph Sambrook and David W. Russel (2001). *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. ISBN 0-87969-576-5. Chapter 8: In vitro Amplification of DNA by the Polymerase Chain Reaction
4. ^ Pavlov AR, Pavlova NV, Kozyavkin SA, Slesarev AI (2004). "Recent developments in the optimization of thermostable DNA polymerases for efficient applications". *Trends Biotechnol.* **22**: 253-260. PMID 15109812.
5. ^ Rychlik W, Spencer WJ, Rhoads RE (1990). "Optimization of the annealing temperature for DNA amplification in vitro". *Nucl Acids Res* **18**: 6409-6412.
6. ^ [a] [b] D.J. Sharkey, E.R. Scalice, K.G. Christy Jr., S.M. Atwood, and J.L. Daiss (1994). "Antibodies as Thermolabile Switches: High Temperature Triggering for the Polymerase Chain Reaction". *Bio/Technology* **12**: 506-509.
7. ^ Chien A, Edgar DB, Trela JM (1976). "Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus". *J. Bacteriol* **174**: 1550-1557. PMID 8432.
8. ^ Lawyer FC, Stoffel S, Saiki RK, Chang SY, Landre PA, Abramson RD, Gelfand DH (1993). "High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity". *PCR Methods Appl.* **2**: 275-287. PMID 8324500.
9. ^ Chemical Synthesis, Sequencing, and Amplification of DNA (class notes on MBB/BIO 343). Arizona State University. Retrieved on 29 October 2007.
10. ^ Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986). "Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes.". *Nature* **324**: 163-166. PMID 3785382.
11. ^ Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL (1995). "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides". *Gene* **164**: 49-53. PMID 7590320.
12. ^ Innis MA, Myambo KB, Gelfand DH, Brow MA. (1988). "DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA". *Proc Natl Acad Sci USA* **85**: 9436-4940. PMID 3200828.
13. ^ Pierce KE and Wangh LJ (2007). "Linear-after-the-exponential polymerase chain reaction and allied technologies Real-time detection strategies for rapid, reliable diagnosis from single cells". *Methods Mol Med.* **132**: 65-85. PMID 17876077.
14. ^ Pavlov AR, Pavlova NV, Kozyavkin SA, Slesarev AI (2006). "Thermostable DNA Polymerases for a Wide Spectrum of Applications: Comparison of a Robust Hybrid TopoTaq to other enzymes", in Kieleczawa J: *DNA Sequencing II: Optimizing Preparation and Cleanup*. Jones and Bartlett, pp. 241-257. ISBN 0-7637338-3-0.
15. ^ Myriam Vincent, Yan Xu and Huimin Kong (2004). "Helicase-dependent isothermal DNA amplification". *EMBO reports* **5** (8): 795–800.
16. ^ Q. Chou, M. Russell, D.E. Birch, J. Raymond and W. Bloch (1992). "Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications". *Nucleic Acids Research* **20**: 1717-1723.
17. ^ E. Zietkiewicz, A. Rafalski, and D. Labuda (1994). "Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification". *Genomics* **20** (2): 176-83.
18. ^ Ochman H, Gerber AS, Hartl DL (1988). "Genetic applications of an inverse polymerase chain reaction". *Genetics* **120**: 621-623. PMID 2852134.
19. ^ Mueller PR, Wold B (1988). "In vivo footprinting of a muscle specific enhancer by ligation mediated PCR". *Science* **246**: 780-786. PMID 2814500.
20. ^ Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB (1996). "Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands". *Proc Natl Acad Sci U S A* **93** (13): 9821-9826. PMID 8790415.
21. ^ Y.G. Liu and R. F. Whittier (1995). "Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking". *Genomics* **25** (3): 674-81.
22. ^ Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991). "'Touchdown' PCR to circumvent spurious priming during gene amplification.". *Nucl Acids Res* **19**: 4008.
23. ^ David, F.Turlotte, E., (1998). "An Isothermal Amplification Method". *C.R.Acad. Sci Paris, Life Science* **321** (1): 909-914.
24. ^ Fabrice David (September-October 2002). Utiliser les propriétés topologiques de l'ADN: une nouvelle arme contre les agents pathogènes. Fusion.(in French)
25. ^ Kleppe K, Ohtsuka E, Kleppe R, Molineux I, Khorana HG (1971). "Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases". *J. Mol. Biol.* **56**: 341-361.
26. ^ [a] [b] Mullis, Kary (1998). *Dancing Naked in the Mind Field*. New York: Pantheon Books. ISBN 0-679-44255-3.
27. ^ Rabinow, Paul (1996). *Making PCR: A Story of Biotechnology*. Chicago: University of Chicago Press. ISBN 0-226-70146-8.

28. ^ Mullis, Kary (1990). "The unusual origin of the polymerase chain reaction". *Scientific American* **262** (4): 56-61, 64-5.
29. ^ Advice on How to Survive the Taq Wars ¶2: GEN Genetic Engineering News Biobusiness Channel: Article. May 1 2006 (Vol. 26, No. 9).

DNA polymerase

From Wikipedia, the free encyclopedia

• *Have questions? Find out how to ask questions and get answers.* •

3D structure of the DNA-binding helix-hairpin-helix motifs in human DNA polymerase beta

A DNA polymerase is an enzyme that assists in DNA replication. Such enzymes catalyze the polymerization of deoxyribonucleotides alongside a DNA strand, which they "read" and use as a template. The newly-polymerized molecule is complementary to the template strand and identical to the template's partner strand.

DNA polymerase is considered to be a holoenzyme since it requires a magnesium ion as a co-factor to function properly. In the absence of the magnesium ion, it is referred to as an apoenzyme.

DNA-Polymerase initiates DNA replication by binding to a piece of single-stranded DNA.

Contents

- 1 Function
- 2 Variation across species
- 3 DNA polymerase families
  - 3.1 Family A
  - 3.2 Family B
  - 3.3 Family C
  - 3.4 Family D
  - 3.5 Families X
  - 3.6 Families Y
  - 3.7 Family RT
- 4 Prokaryotic DNA polymerases
- 5 Eukaryotic DNA polymerases
- 6 See also
- 7 References
  - 7.1 Citations
- 8 External links

Function

DNA replication

DNA polymerase can only add free nucleotides to the 3’ end of the newly forming strand. This results in elongation of the new strand in a 5'-3' direction. No known DNA polymerase is able to begin a new chain (*de novo*). They can only add a nucleotide onto a preexisting 3'-OH group. For this reason, DNA polymerase needs a primer at which it can add the first nucleotide. Primers consist of RNA and DNA bases with the first

two bases always being RNA, and are synthesized by another enzyme called primase. An enzyme known as a helicase is required to unwind DNA from a double-strand structure to a single-strand structure to facilitate replication of each strand consistent with the semiconservative model of DNA replication.

Error correction is a property of some, but not all, DNA polymerases. This process corrects mistakes in newly-synthesized DNA. When an incorrect base pair is recognized, DNA polymerase reverses its direction by one base pair of DNA. The 3'->5' exonuclease activity of the enzyme allows the incorrect base pair to be excised (this activity is known as *proofreading*). Following base excision, the polymerase can re-insert the correct base and replication can continue.

Variation across species

DNA polymerases have highly-conserved structure, which means that their overall catalytic subunits vary, on a whole, very little from species to species. Conserved structures usually indicate important, irreplacable functions of the cell, the maintenance of which provides evolutionary advantages.

Some viruses also encode special DNA polymerases which may selectively replicate viral DNA through a variety of mechanisms. Retroviruses encode an unusual DNA polymerase called reverse transcriptase, which is an RNA-dependent DNA polymerase (RdDp). It polymerizes DNA from a template of RNA.

DNA polymerase families

Based on sequence homology, DNA polymerases can be further subdivided into seven different families A, B, C, D, X, Y, and RT.

Family A

Family A polymerases contain both replicative and repair polymerases. Replicative members from this family include the extensively studied T7 DNA polymerase as well as the eukaryotic mitochondrial DNA Polymerase γ. Among the repair polymerases are *E. coli* DNA pol I, *Thermus aquaticus* pol I, and *Bacillus stearothermophilus* pol I. These repair polymerases are involved in excision repair and processing of Okazaki fragments generated during lagging strand synthesis.

Family B

Family B polymerases mostly contain replicative polymerases and include the major eukaryotic DNA polymerases α, δ, ε, (see Greek letters used in mathematics) and also DNA polymerase ζ. Family B also includes DNA polymerases encoded by some bacteria and bacteriophages, of which the best characterized are from T4, Phi29 and RB69 bacteriophages. These enzymes are involved in both leading and lagging strand synthesis. A hallmark of the B family of polymerases is remarkable accuracy during replication and many have strong 3'-5' exonuclease activity (except DNA polymerase α and ζ which have no proofreading activity).

Family C

Family C polymerases are the primary bacterial chromosomal replicative enzymes. DNA Polymerase III alpha subunit from *E. coli* possesses no known nuclease activity. A separate subunit, the epsilon subunit, possesses the 3'-5' exonuclease activity used for editing during chromosomal replication.

Family D

Family D polymerases are still not very well characterized. All known examples are found in the Euryarchaeota subdomain of Archaea and are thought to be replicative polymerases.

Families X

Family X contains the well known eukaryotic polymerase pol β as well as other eukaryotic polymerases such as pol σ, pol λ, pol μ, and terminal deoxynucleotidyl transferase (TdT). Pol β is required for short-patch base excision repair, a DNA repair pathway that is essential for repairing abasic sites. Pol λ and Pol μ are involved in non-homologous end joining, a mechanism for rejoining DNA double-strand breaks. TdT is only expressed in lymphoid tissue and adds "n nucleotides" to double-strand breaks formed during V(D)J recombination to promote immunological diversity. The yeast *Saccharomyces cerevisiae* has only one Pol X polymerase, Pol4, which is involved in non-homologous end joining.

Families Y

The Y-family polymerases differ from others in having a low fidelity on undamaged templates and in their ability to replicate through damaged DNA. Members of this family are hence called translesion sythesis (TLS) polymerases. Depending on the lesion TLS polymerases can bypass the damage in an error-free or error-prone fashion, the latter resulting in elevated mutagenesis. Xeroderma pigmentosum variant (XPV) patients for instance have mutations in the gene encoding Pol η (eta), which is error-free for UV-lesions. In XPV patients alternative error-prone polymerases e.g. Polζ (zeta) (polymerase ζ is a B Family polymerase), are thought to be involved in mistakes which result in the cancer predisposition of these patients. Other members in humans are Pol ι (iota), Pol κ (kappa) and Rev1 (terminal deoxycytidyl transferase). In E.coli two TLS polymerases, Pol IV (DINB) and PolV (UMUC), are known.

Family RT

Finally, the reverse transcriptase family contain examples both from retroviruses and eukaryotic polymerases. The eukaryotic polymerases are usually restricted to telomerases. These polymerases use a RNA template to synthesize the DNA strand.

Prokaryotic DNA polymerases

Bacteria have 5 known DNA polymerases:

- **Pol I**: implicated in DNA repair; has both 5'->3'(Nick translation) and 3'->5' (Proofreading) exonuclease activity.
- **Pol II**: involved in replication of damaged DNA; has both 5'->3'chain extension ability and 3'->5' exonuclease activity.
- **Pol III**: the main polymerase in bacteria (elongates in DNA replication); has 3'->5' exonuclease proofreading ability.
- **Pol IV**: a Y-family DNA polymerase.
- **Pol V**: a Y-family DNA polymerase; participates in bypassing DNA damage.

Eukaryotic DNA polymerases

Eukaryotes have at least 15 DNA Polymerases[1]:

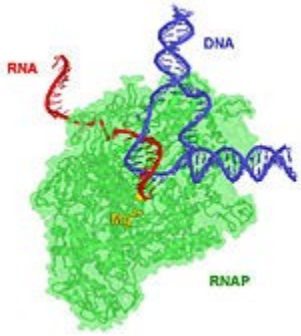
- **Pol α** (synonymes are DNA primase, **RNA** polymerase): acts as a primase (synthesizing a RNA primer), and then as a DNA Pol elongating that primer with DNA nucleotides. After around 20 nucleotides[2] elongation is taken over by Pol δ (on the lagging strand) and ε (on the leading strand).
- **Pol β**: is implicated in repairing DNA.
- **Pol γ**: replicates mitochondrial DNA.
- **Pol δ**: is the main polymerase on the lagging strand in eukaryotes, it is highly processive and has 3'->5' exonuclease activity.
- **Pol ε**: is the primary leading strand DNA polymerase in eukaryotes, and is also highly processive and has 3'->5' exonuclease activity [3].
- **η**, **ι**, **κ**, and **Rev1** are Y-family DNA polymerases and **Pol ζ** is a B-family DNA polymerase. These polymerases are involved in the bypass of DNA damage.[4]
- There are also other eukaryotic polymerases known, which are not as well characterized: **θ**, **λ**, **φ**, **σ**, and **μ**. There are also others, but the nomenclature has become quite jumbled.

None of the eukariotic polymerases can remove primers (5'->3' exonuclease activity), that function is carried out by other enzymes. Only the polymerases that deal with the elongation (γ, δ and ε) have proofreading ability (3'->5' exonuclease).

RNA polymerase

From Wikipedia, the free encyclopedia

• *Have questions? Find out how to ask questions and get answers.* •



RNAP from *T. aquaticus* pictured during elongation. Portions of the enzyme were made transparent so as to make the path of RNA and DNA more clear. The magnesium ion (yellow) is located at the enzyme active site.

RNA polymerase (**RNAP** or **RNApol**) is an enzyme that makes an RNA copy of a DNA or RNA template. In cells, RNAP is needed for constructing RNA chains from DNA genes, a process called transcription. RNA polymerase enzymes are essential to life and are found in all organisms and many viruses. In chemical terms, RNAP is a nucleotidyl transferase that polymerizes ribonucleotides at the 3' end of an RNA transcript.

Contents

- 1 History
- 2 Control of transcription
- 3 RNA polymerase action
  - 3.1 Binding and initiation
  - 3.2 Elongation
  - 3.3 Termination
- 4 RNA polymerase in bacteria

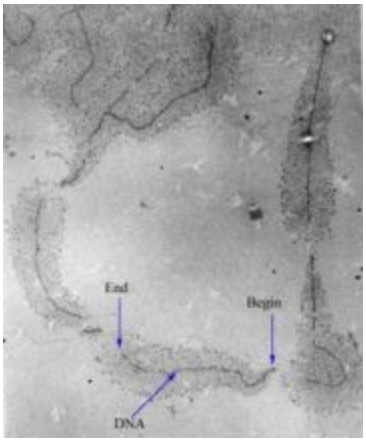
- 4.1 Transcriptional cofactors
- 5 RNA polymerase in eukaryotes
- 6 RNA polymerase in archaea
- 7 RNA polymerase in viruses
- 8 Functional domains
  - 8.1 C-terminal domain of RNA polymerase
    - 8.1.1 Transcription Initiation
    - 8.1.2 5'Capping
    - 8.1.3 Spliceosome
    - 8.1.4 Mutation in the CTD
- 9 RNA polymerase purification
- 10 See also
- 11 References

- 12 External links

History

RNAP was discovered independently by Sam Weiss and Jerard Hurwitz in 1960.[1] By this time the 1959 Nobel Prize in Medicine had been awarded to Severo Ochoa and Arthur Kornberg for the discovery of what was believed to be RNAP[2], but instead turned out to be a ribonuclease.
The 2006 Nobel Prize in Chemistry was awarded to Roger Kornberg for creating detailed molecular images of RNA polymerase during various stages of the transcription process.[3]

Control of transcription



An electron-micrograph of DNA strands decorated by hundreds of RNAP molecules too small to be resolved. Each RNAP is transcribing an RNA strand which can be seen branching off of the DNA. "Begin" indicates the 3' end of the DNA, where RNAP initiates transcription; "End" indicates the 5' end, where the longer RNA molecules are almost completely transcribed.
Control of the process of gene transcription affects patterns of gene expression and thereby allows a cell to adapt to a changing environment, perform specialized roles within an organism, and maintain basic metabolic processes necessary for survival. Therefore, it is hardly surprising that the activity of RNAP is both complex and highly regulated. In *Escherichia coli* bacteria, more than 100 factors have been identified which modify the activity of RNAP.[4]

RNAP can initiate transcription at specific DNA sequences known as promoters. It then produces an RNA chain which is complementary to the template DNA strand. The process of adding nucleotides to the RNA

strand is known as elongation; In eukaryotes, RNAP can build chains as long as 2.4 million nucleosides (the full length of the dystrophin gene). RNAP will preferentially release its RNA transcript at specific DNA sequences encoded at the end of genes known as terminators.

Products of RNAP include:

- Messenger RNA (mRNA)—template for the synthesis of proteins by ribosomes.
- Non-coding RNA or "RNA genes"—a broad class of genes that encode RNA that is not translated into protein. The most prominent examples of RNA genes are transfer RNA (tRNA) and ribosomal RNA (rRNA), both of which are involved in the process of translation. However, since the late 1990s, many new RNA genes have been found, and thus RNA genes may play a much more significant role than previously thought.
    - Transfer RNA (tRNA)—transfers specific amino acids to growing polypeptide chains at the ribosomal site of protein synthesis during translation
    - Ribosomal RNA (rRNA)—a component of ribosomes
    - Micro RNA—regulates gene activity
    - Catalytic RNA (Ribozyme)—enzymatically active RNA molecules

RNAP accomplishes *de novo* synthesis. It is able to do this because specific interactions with the initiating nucleotide hold RNAP rigidly in place, facilitating chemical attack on the incoming nucleotide. Such specific interactions explain why RNAP prefers to start transcripts with ATP (followed by GTP, UTP, and then CTP). In contrast to DNA polymerase, RNAP includes helicase activity, therefore no separate enzyme is needed to unwind DNA.

RNA polymerase action

Binding and initiation

RNA Polymerase binding involves the α subunit recognizing the upstream element (-40 to -70 base pairs) in DNA, as well as the σ factor recognizing the -10 to -35 region. There are numerous σ factors that regulate gene expression. For example, σ^{70} is expressed under normal conditions and allows RNAP binding to house-keeping genes, while σ^{32} elicits RNAP binding to heat-shock genes.

After binding to the DNA, the RNA polymerase switches from a closed complex to an open complex. This change involves the separation of the DNA strands to form a unwound section of DNA of approximately 13bp. Ribonucleotides are base-paired to the template DNA strand, according to Watson-Crick base-pairing interactions. Supercoiling plays an important part in polymerase activity because of the unwinding and rewinding of DNA. Because regions of DNA in front of RNAP are unwound, there is compensatory positive supercoils. Regions behind RNAP are rewound and negative supercoils are present.

Elongation

Transcription elongation involves the further addition of ribonucleotides and the change of the open complex to the transcriptional complex. RNAP cannot start forming full length transcripts because of its strong binding to promoter. Transcription at this stage primarily results in short RNA fragments of around 9 bp in a process known as abortive transcription. Once the RNAP starts forming longer transcripts it clears the promoter. At this point, the -10 to -35 promoter region is disrupted, and the σ factor falls off RNAP. This allows the rest of the RNAP complex to move forward, as the σ factor held the RNAP complex in place.

The 17 bp transcriptional complex has an 8 bp DNA-RNA hybrid, that is, 8 base-pairs involve the RNA transcript bound to the DNA template strand. As transcription progresses, ribonucleotides are added to the 3' end of the RNA transcript and the RNAP complex moves along the DNA. Although RNAP does not seem to have the 3'exonuclease activity that characterizes the *proofreading* activity found in DNA polymerase, there is evidence of that RNAP will halt at mismatched base-pairs and correct it.

The addition of ribonucleotides to the RNA transcript has a very similar mechanism to DNA polymerization - it is believed that these polymerases are evolutionarily related. Aspartyl (asp) residues in the RNAP will hold onto Mg^{2+} ions, which will in turn coordinate the phosphates of the ribonucleotides. The first Mg^{2+} will hold onto the α-phosphate of the NTP to be added. This allows the nucleophilic attack of the 3'OH from the RNA transcript, adding an additional NTP to the chain. The second Mg^{2+} will hold onto the pyrophosphate of the NTP. The overall reaction equation is:

$(NMP)_n + NTP \rightarrow (NMP)_{n+1} + PP_i$

Termination

Termination of RNA transcription can be rho-independent or rho-dependent:

Rho-independent transcription termination is the termination of transcription without the aid of the rho protein. Transcription of a palindromic region of DNA causes the formation of a *hairpin* structure from the RNA transcription looping and binding upon itself. This hairpin structure is often rich in G-C base-pairs, making it more stable than the DNA-RNA hybrid itself. As a result, the 8bp DNA-RNA hybrid in the transcription complex shifts to a 4bp hybrid. Coincidentally, these last 4 base-pairs are weak A-U base-pairs, and the entire RNA transcript will fall off.[5]

RNA polymerase in bacteria

In bacteria, the same enzyme catalyzes the synthesis of mRNA and ncRNA.

RNAP is a relatively large molecule. The core enzyme has 5 subunits (~400 kDa):

- α_2: the two α subunits assemble the enzyme and recognize regulatory factors. Each subunit has two domains: αCTD (C-Terminal domain) binds the UP element of the extended promoter, and αNTD (N-terminal domain) binds the rest of the polymerase.
- β: this has the polymerase activity (catalyzes the synthesis of RNA) which includes chain initiation and elongation.
- β': binds to DNA (nonspecifically).
- ω: restores denatured RNA polymerase to its functional form in vitro. It has been observed to offer a protective/chaperone function to the β' subunit in *Mycobacterium smegmatis*. Now known to promote assembly.

In order to bind promoter-specific regions, the core enzyme requires another subunit, sigma (σ). The sigma factor greatly reduces the affinity of RNAP for nonspecific DNA while increasing specificity for certain promoter regions, depending on the sigma factor. That way, transcription is initiated at the right region. The complete holoenzyme therefore has 6 subunits: $\alpha_2\beta\beta'\sigma\omega$ (~480 kDa). The structure of RNAP exhibits a groove with a length of 55 Å (5.5 nm) and a diameter of 25 Å (2.5 nm). This groove fits well the 20 Å (2 nm) double strand of DNA. The 55 Å (5.5 nm) length can accept 16 nucleotides.

When not in use RNA polymerase binds to low affinity sites to allow rapid exchange for an active promoter site when one opens. RNA polymerase holoenzyme, therefore, does not freely float around in the cell when not in use.

Transcriptional cofactors

There are a number of proteins which can bind to RNAP and modify its behavior. For instance, greA and greB from *E. coli* can enhance the ability of RNAP to cleave the RNA template near the growing end of the chain. This cleavage can rescue a stalled polymerase molecule, and is likely involved in proofreading the occasional mistakes made by RNAP. A separate cofactor, Mfd, is involved in transcription-coupled repair, the process in which RNAP recognizes damaged bases in the DNA template and recruits enzymes to restore the DNA. Other cofactors are known to play regulatory roles, i.e. they help RNAP choose whether or not to express certain genes.

RNA polymerase in eukaryotes

Essential Subunit Of Human RNA Polymerases I, II and III

Eukaryotes have several types of RNAP, characterized by the type of RNA they synthesize:

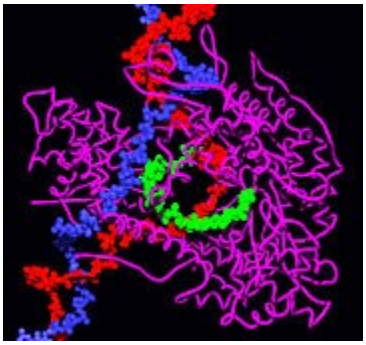
- RNA polymerase I synthesizes a pre-rRNA 45S, which matures into 28S, 18S and 5.8S rRNAs which will form the major RNA sections of the ribosome.[6]
- RNA polymerase II synthesizes precursors of mRNAs and most snRNA and microRNAs.[7] This is the most studied type, and due to the high level of control required over transcription a range of transcription factors are required for its binding to promoters.
- RNA polymerase III synthesizes tRNAs, rRNA 5S and other small RNAs found in the nucleus and cytosol.[8]

There are other RNA polymerase types in mitochondria and chloroplasts.

RNA polymerase in archaea

Archaea have a single RNAP that is closely related to the three main eukaryotic polymerases. Thus, it has been speculated that the archaeal polymerase resembles the ancestor of the specialized eukaryotic polymerases.[9]

RNA polymerase in viruses



T7 RNA polymerase producing a mRNA (green) from a DNA template. The protein is shown as a purple ribbon. Image derived from PDB 1MSW.
Many viruses also encode for RNAP. Perhaps the most widely studied viral RNAP is found in bacteriophage T7. This single-subunit RNAP is related to that found in mitochondria and chloroplasts, and shares considerable homology to DNA polymerase.[10] It is believed that most viral polymerases therefore evolved from DNA polymerase and are not directly related to the multi-subunit polymerases described above.
The viral polymerases are diverse, and include some forms which can use RNA as a template instead of DNA. This occurs in negative strand RNA viruses and dsRNA viruses, both of which exist for a portion of their life cycle as double-stranded RNA. However, some positive strand RNA viruses, such as polio, also contain these RNA dependent RNA polymerases.[11]

Functional domains

C-terminal domain of RNA polymerase

Transcription Initiation

The carboxy-terminal domain (CTD) of **RNA polymerase** II is that portion of the polymerase which is involved in the initiation of DNA transcription. The CTD typically consists of up to 52 repeats of the sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser [12]. The transcription factor TFIIH is a kinase and will hyperphosphorylate the CTD of RNAP, and in doing so, causes the RNAP complex to move away from the initiation site.

5'Capping

The carboxy-terminal domain is also the binding site of the cap-synthesizing and cap-binding complex. In eukaryotes, after transcription of the 5' end of an RNA transcript, the cap-synthesizing complex on the CTD will remove the gamma-phosphate from the 5'phosphate and attach a GMP, forming a 5',5'-triphosphate linkage. The synthesizing complex falls off and the cap then binds to the cap-binding complex (CBC), which is bound to the CTD.

The 5'cap of eukaryotic RNA transcripts is important for binding of the RNA transcript to the ribosome during translation, to the CTD of RNAP, and prevents RNA degradation.

Spliceosome

The carboxy-terminal domain is also the binding site for spliceosome factors that are part of RNA splicing. These allow for the splicing and removal of introns (in the form of a lariat structure) during RNA transcription.

Mutation in the CTD

Major studies have been carried out in which knockout of particular amino acids was achieved in the CTD. The results indicate that RNA polymerase II CTD truncation mutations affect the ability to induce transcription of a subset of genes *in vivo*, and the lack of response to induction maps to the upstream activating sequences of these genes.

RNA polymerase purification

RNA polymerase can be isolated in the following ways:

- By a phosphocellulose column.[13]
- By glycerol gradient centrifugation.[14]
- By a DNA column.
- By an Ion exchange column.[15]

And also combinations of the above techniques.

See also

- DNA polymerase
- T7 RNA polymerase
- Alpha-amanitin

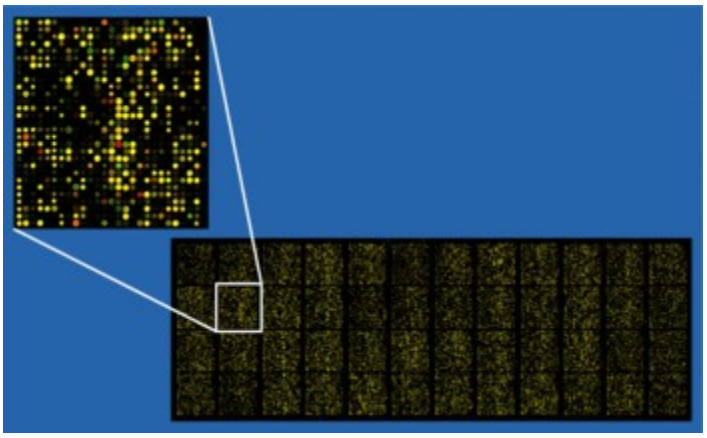
References

1. ^ Jerard Hurwitz (Dec 2005). "The Discovery of RNA Polymerase". *Journal of Biological Chemistry* **280** (52): 42477-85. doi:10.1074/jbc.X500006200. PMID 16230341.
2. ^ Nobel Prize 1959
3. ^ Nobel Prize in Chemistry 2006
4. ^ Akira Ishihama (2000). "Functional modulation of Escherichia coli RNA polymerase" **54**: 499-518. PMID 11018136.
5. ^ Farnham PJ; Platt T. (Feb 1981). "Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro". *Nucleic Acids Res.* **9** (3): 563-77. PMID 7012794.
6. ^ Grummt I. (1999). "Regulation of mammalian ribosomal gene transcription by RNA polymerase I.". *Prog Nucleic Acid Res Mol Biol.* **62**: 109-54. PMID 9932453.
7. ^ Lee Y; Kim M; Han J; Yeom KH; Lee S; Baek SH; Kim VN. (Oct 2004). "MicroRNA genes are transcribed by RNA polymerase II". *EMBO J.* **23** (20): 4051-60. PMID 15372072.
8. ^ Willis IM. (Feb 1993). "RNA polymerase III. Genes, factors and transcriptional specificity". *Eur J Biochem.* **212** (1): 1-11. PMID 8444147.
9. ^ D Langer, J Hain, P Thuriaux and W Zillig (1995) Transcription in Archaea: Similarity to that in Eucarya PNAS **92** 5768-5772
10. ^ Hedtke *et al.* (1997) Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. Science **227** 809-811
11. ^ Paul Ahlquist (2002) RNA-Dependent RNA Polymerases, Viruses, and RNA Silencing. Science **296** 1270-1273
12. ^ Anton Meinhart1; Patrick Cramer (Jul 2004). "Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors". *Nature* **430** (6996): 223-226. doi:10.1038/nature02679. PMID 15241417.
13. ^ Kelly JL; Lehman IR. (Aug 1986). "Yeast mitochondrial RNA polymerase. Purification and properties of the catalytic subunit.". *J Biol Chem.* **261** (22): 10340-7. PMID 3525543.
14. ^ Honda A et al (Apr 1990). "Purification and molecular structure of RNA polymerase from influenza virus A/PR8.". *J Biochem (Tokyo)* **107** (4): 624-8. PMID 2358436.
15. ^ Hager *et al.* (1990) Use of Mono Q High-Resolution Ion-Exchange Chromatography To Obtain Highly Pure and Active Escherichia coli RNA Polymerase Biochemistry **29** 7890-7894

DNA microarray

From Wikipedia, the free encyclopedia

• Interested in contributing to Wikipedia? •


Example of an approximately 40,000 probe spotted oligo microarray with enlarged inset to show detail.

A **DNA microarray** (also commonly known as *gene or genome chip*, *DNA chip*, or *gene array*) is a collection of microscopic DNA spots, commonly representing single genes, arrayed on a solid surface by covalent attachment to a chemical matrix. DNA arrays are different from other types of *microarray* only in that they either measure DNA or use DNA as part of its detection system. Qualitative or quantitative measurements with DNA microarrays utilize the selective nature of DNA-DNA or DNA-RNA hybridization under high-stringency conditions and fluorophore-based detection. DNA arrays are commonly used for expression profiling, i.e., monitoring expression levels of thousands of genes simultaneously, or for comparative genomic hybridization.

Contents

- 1 Introduction
- 2 Fabrication
  - 2.1 Spotted vs. Oligonucleotide Arrays
  - 2.2 Two-color vs. one-color detection
- 3 Genotyping microarrays
  - 3.1 Standardization
  - 3.2 Statistical analysis
  - 3.3 Relation between probe and gene
- 4 Public databases of microarray data
- 5 Online microarray data analysis programs and tools
- 6 Notable microarray related articles
- 7 References

- 8 External links

Introduction

Arrays of DNA can either be spatially arranged, as in the commonly known *gene or genome chip*, *DNA chip*, or *gene array*, or can be specific DNA sequences tagged or labelled such that they can be independently identified in solution. The traditional solid-phase array is a collection of microscopic DNA spots attached to a solid surface, such as glass, plastic or silicon chip. The affixed DNA segments are known as *probes* (although some sources will use different nomenclature such as *reporters*), thousands of which can be placed in known locations on a single DNA microarray. Microarray technology evolved from Southern blotting, whereby fragmented DNA is attached to a substrate and then probed with a known gene or fragment. DNA microarrays can be used to detect DNA (e.g., in comparative genomic hybridization); it also permits detection of RNA

(most commonly as cDNA after reverse transcription) that may or may not be translated into proteins, which is referred to as "expression analysis" or expression profiling.

Since there can be tens of thousands of distinct probes on an array, each microarray experiment can potentially accomplish the equivalent number of genetic tests in parallel. Arrays have therefore dramatically accelerated many types of investigations. The use of a collection of distinct DNAs in arrays for expression profiling was first described in 1987, and the arrayed DNAs were used to identify genes whose expression is modulated by interferon. [1] These early gene arrays were made by spotting cDNAs onto filter paper with a pin-spotting device. The use of miniaturized microarrays for gene expression profiling was first reported in 1995, [2] and a complete eukaryotic genome (*Saccharomyces cerevisiae*) on a microarray was published in 1997. [3]

Applications of these arrays include:

- mRNA or gene expression profiling - Monitoring expression levels for thousands of genes simultaneously to study the effects of certain treatments, diseases, and developmental stages on gene expression. For example, microarray-based gene expression profiling can be used to identify disease genes by comparing gene expression in diseased and normal cells.
- Comparative genomic hybridization - Assessing genome content in different cells or closely related organisms. [4] [5]
- SNP detection arrays - Identifying single nucleotide polymorphism among alleles within or between populations.[6]
- Chromatin immunoprecipitation (chIP) studies - Determining protein binding site occupancy throughout the genome, employing ChIP-on-chip technology.

Fabrication

Microarrays can by manufactured in different ways, depending on the number of probes under examination, costs, customization requirements, and the type of scientific question being asked. Arrays may have as few as 10 probes to up to 390,000 micron-scale probes from commercial vendors.

Spotted vs. Oligonucleotide Arrays

Microarrays can be fabricated using a variety of technologies, including printing with fine-pointed pins onto glass slides, photolithography using pre-made masks, photolithography using dynamic micromirror devices, ink-jet printing, [7] or electrochemistry on microelectrode arrays.

A DNA microarray being created

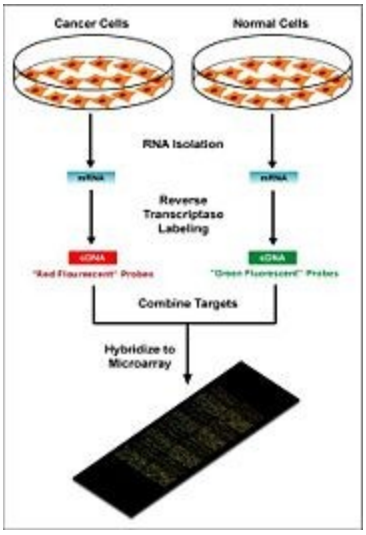
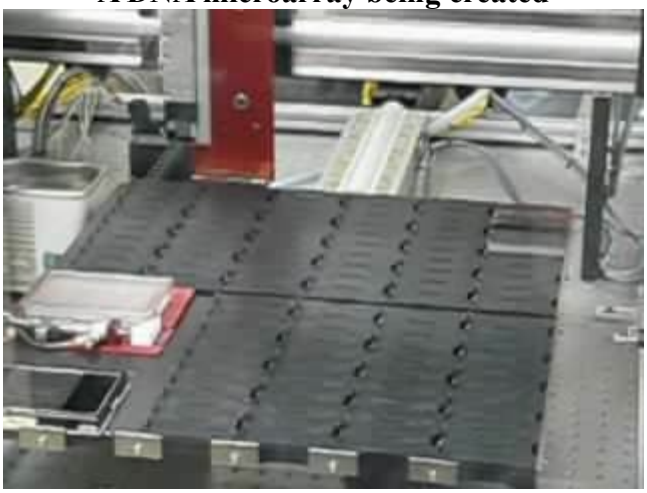


Diagram of typical dual-colour microarray experiment.



A DNA microarray being printed by a robot at the University of Delaware.

In *spotted microarrays*, the probes are oligonucleotides, cDNA or small fragments of PCR products that correspond to mRNAs. There probes are synthesized prior to deposition on the array surface and are then "spotted" onto glass. A common approach utilizes an array of fine pins or needles controlled by a robotic arm that is dipped into wells containing DNA probes and then depositing each probe at designated locations on the array surface. The resulting "grid" of probes represents the nucleic acid profiles of the prepared probes and is ready to receive complementary cDNA or cRNA "targets" derived from experimental or clinical samples.

This technique is used by research scientists around the world to produce "in-house" printed microarrays from their own labs. These arrays may be easily customized for each experiment, because researchers can choose the probes and printing locations on the arrays, synthesize the probes in their own lab (or collaborating facility), and spot the arrays. They can then generate their own labeled samples for hybridization, hybridize the samples to the array, and finally scan the arrays with their own equipment. This provides a relatively low-cost microarray that is customized for each study, and avoids the costs of purchasing often more expensive commercial arrays that may represent vast numbers of genes that are not of interest to the investigator.

Publications exist which indicate in-house spotted microarrays may not provide the same level of sensitivity compared to commercial oligonucleotide arrays, [8] possibly owing to the small batch sizes and reduced printing efficiencies when compared to industrial manufactures of oligo arrays. GE Healthcare offers a commercial array platform called the "Code Link" system where 30-mer oligonucleotide probes (sequences of 30 nucleotides in length) are piezoelectrically deposited on an acrylamide matrix without any contact being made between the depositing equipment and the array surface itself. These arrays are comparable in quality to most manufactures arrays and generally superior to in-house printed arrays.

In *oligonucleotide microarrays*, the probes are short sequences designed to match parts of the sequence of known or predicted open reading frames. Although oligonucleotide probes are often used in "spotted" microarrays, the term "oligonucleotide array" most often refers to a specific technique of manufacturing. Oligonucleotide arrays are produced by printing short oligonucleotide sequences designed to represent a single gene or family of gene splice-variants by synthesizing this sequence directly onto the array surface instead of depositing intact sequences. Sequences may be longer (60-mer probes such as the Agilent design) or shorter (25-mer probes produced by Affymetrix) depending on the desired purpose; longer probes are more specific to individual target genes, shorter probes may be spotted in higher density across the array and are cheaper to manufacture.

One technique used to produce oligonucleotide arrays include photolithographic synthesis (Agilent and

Affymetrix) on a silica substrate where light and light-sensitive masking agents are used to "build" a sequence one nucleotide at a time across the entire array. [9] Each applicable probe is selectively "unmasked" prior to bathing the array in a solution of a single nucleotide, then a masking reaction takes place and the next set of probes are unmasked in preparation for a different nucleotide exposure. After many repetitions, the sequences of every probe become fully constructed. More recently, Maskless Array Synthesis from NimbleGen Systems has combined flexibility with large numbers of probes. [10]

Two-color vs. one-color detection

Two-Color microarrays are typically hybridized with cDNA prepared from two samples to be compared (e.g. diseased tissue versus healthy tissue) and that are labeled with two different fluorophores. [11] Fluorescent dyes commonly used for cDNA labelling include Cy3, which has a fluorescence emission wavelength of 570 nm (corresponding to the green part of the light spectrum), and Cy5 with a fluorescence emission wavelength of 670 nm (corresponding to the red part of the light spectrum). The two Cy-labelled cDNA samples are mixed and hybridized to a single microarray that is then scanned in a microarray scanner to visualize fluorescence of the two fluorophores after excitation with a laser beam of a defined wavelength. Relative intensities of each fluorophore may then be used in ratio-based analysis to identify up-regulated and down-regulated genes. [12]

Oligonucleotide microarrays often contain control probes designed to hybridize with RNA spike-ins. The degree of hybridization between the spike-ins and the control probes is used to normalize the hybridization measurements for the target probes. Although absolute levels of gene expression may be determined in the two-color array, the relative differences in expression among different spots within a sample and between samples is the preferred method of data analysis for the two-color system. Examples of providers for such microarrays includes Agilent with their Dual-Mode platform, Eppendorf with their DualChip platform, and TeleChem International with ArrayIt.

Two Affymetrix chips

In *single-channel microarrays* or *one-color microarrays*, the arrays are designed to give estimations of the absolute levels of gene expression. Therefore the comparison of two conditions requires two separate single-dye hybridizations. As only a single dye is used, the data collected represent absolute values of gene expression. These may be compared to other genes within a sample or to reference "normalizing" probes used to calibrate data across the entire array and across multiple arrays. Two popular single-channel systems are the Affymetrix "Gene Chip" and GE Healthcare "Code Link" arrays. One strength of the single-dye system lies in the fact that an aberrant sample cannot affect the raw data derived from other samples, because each array chip is exposed to only one sample (as opposed to a two-color system in which a single low-quality sample may drastically impinge on overall data precision even if the other sample was of high quality). Another benefit is that data are more easily compared to arrays from different experiments; the absolute values of gene expression may be compared between studies conducted months or years apart. A drawback to the one-color system is that, when compared to the two-color system, twice as many microarrays are needed to compare samples within an experiment.

Genotyping microarrays

Main article: SNP array

DNA microarrays can also be used to scan the entire sequence of a genome to identify genetic variation at certain locations.

SNP microarrays are a type of DNA microarray that are used to identify genetic variation in individuals and across populations. [6]

Standardization

The lack of standardization in arrays presents an interoperability problem in bioinformatics, which hinders the exchange of array data. Various grass-roots open-source projects are attempting to facilitate the exchange and analysis of data produced with non-proprietary chips.

- The "Minimum Information About a Microarray Experiment" (MIAME) checklist helps define the level of detail that should exist and is being adopted by many journals as a requirement for the submission of papers incorporating microarray results. MIAME describes the minimum required information for complying experiments, but not its format. Thus, as of 2007, whilst many formats can support the MIAME requirements there is no format which permits verification of complete semantic compliance.

- The "MicroArray Quality Control (MAQC) Project" is being conducted by the FDA to develop standards and quality control metrics which will eventually allow the use of MicroArray data in drug discovery, clinical practice and regulatory decision-making. [13]

- The MicroArray and Gene Expression (MAGE) group is working on the standardization of the representation of gene expression data and relevant annotations.

Statistical analysis

The analysis of DNA microarrays poses a large number of statistical problems, including the normalization of the data. There are dozens of proposed normalization methods in the published literature; as in many other cases where authorities disagree, a sound conservative approach is to try a number of popular normalization methods and compare the conclusions reached: how sensitive are the main conclusions to the method chosen?

From a hypothesis-testing perspective, the large number of genes present on a single array means that the experimenter must take into account a multiple testing problem: even if the statistical P-value assigned to a given gene indicates that it is extremely unlikely that differential expression of this gene was due to random rather than treatment effects, the very high number of genes on an array makes it likely that differential expression of some genes represent false positives or false negatives. Statistical methods tailored to microarray analyses have recently become available that assess statistical power based on the variation present in the data and the number of experimental replicates, and can help minimize type I and type II errors in the analyses.[14]

A basic difference between microarray data analysis and much traditional biomedical research is the dimensionality of the data. A large clinical study might collect 100 data items per patient for thousands of patients. A medium-size microarray study will obtain many thousands of numbers per sample for perhaps a hundred samples. Many analysis techniques treat each sample as a single point in a space with thousands of dimensions, then attempt by various techniques to reduce the dimensionality of the data to something humans can visualize.

Relation between probe and gene

The relation between a probe and the mRNA that it is expected to detect is problematic. On the one hand, some mRNAs may cross-hybridize probes in the array that are supposed to detect another mRNA. On the other hand, probes that are designed to detect the mRNA of a particular gene may be relying on genomic EST information that is incorrectly associated with that gene.

Public databases of microarray data

Database	Microarray Experiment Sets	Sample Profiles	as of Date
Gene Expression Omnibus - NCBI	5366	134669	April 1, 2007
Stanford Microarray database	12742	?	April 1, 2007
UPenn RAD database	~100	~2500	Sept. 1, 2007
UNC Microarray database	~31	2093	April 1, 2007
MUSC database	~45	555	April 1, 2007
ArrayExpress at EBI	1643	136	April 1, 2007
caArray at NCI	41	1741	November 15, 2006
UPSC-BASE	~100	?	November 15, 2007

- For a directory of Microarray Databases, see: Gene Expression Databases at the Open Directory Project
- See also the Microarray databases page in Wikipedia

Online microarray data analysis programs and tools

Several Open Directory Project categories list online microarray data analysis programs and tools:

- Bioinformatics : Online Services : Gene Expression and Regulation at the Open Directory Project
- Gene Expression : Databases at the Open Directory Project
- Gene Expression : Software at the Open Directory Project
- Data Mining : Tool Vendors at the Open Directory Project
- Bioconductor: open source and open development software project for the analysis and comprehension of genomic data
- Genevestigator : Web-based database and analysis tool to study gene expression across large sets of tissues, developmental stages, drugs, stimuli, and genetic modifications.

Notable microarray related articles

- Affymetrix
- Applied Microarrays
- Agilent Technologies
- CombiMatrix
- Eppendorf
- Nanogen

    -
    - For microarray companies, see: Products and Services for Gene Expression at the Open Directory Project

References

1. ^ Kulesh DA, Clive DR, Zarlenga DS, Greene JJ (1987). "Identification of interferon-modulated proliferation-related cDNA sequences". *Proc Natl Acad Sci USA* **84**: 8453-8457. PMID 2446323.
2. ^ Schena M, Shalon D, Davis RW, Brown PO (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray". *Science* **270**: 467-470. PMID 7569999.
3. ^ Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW (1997). "Yeast microarrays for genome wide parallel genetic and gene expression analysis". *Proc Natl Acad Sci USA* **94**: 13057-13062. PMID 9371799.
4. ^ Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO (1999). "Genome-wide analysis of DNA copy-number changes using cDNA microarrays". *Nat Genet* **23**: 41-46. PMID 10471496.
5. ^ Moran G, Stokes C, Thewes S, Hube B, Coleman DC, Sullivan D (2004). "Comparative genomics using Candida albicans DNA microarrays reveals absence and divergence of virulence-associated genes in Candida dubliniensis". *Microbiology* **150**: 3363-3382. PMID 15470115.
6. ^ *a* *b* Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SP, Collins FS (1999). "Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays". *Nat Genet* **22**: 164-167. PMID 10369258.
7. ^ http://genomebiology.com/2004/5/8/R58
8. ^ Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, Bumgarner RE, Bushel PR, Chaturvedi K, Choi D, Cunningham ML, Deng S, Dressman HK, Fannin RD, Farin FM, Freedman JH, Fry RC, Harper A, Humble MC, Hurban P, Kavanagh TJ, Kaufmann WK, Kerr KF, Jing L, Lapidus JA, Lasarev MR, Li J, Li YJ, Lobenhofer EK, Lu X, Malek RL, Milton S, Nagalla SR, O'malley JP, Palmer VS, Pattee P, Paules RS, Perou CM, Phillips K, Qin LX, Qiu Y, Quigley SD, Rodland M, Rusyn I, Samson LD, Schwartz DA, Shi Y, Shin JL, Sieber SO, Slifer S, Speer MC, Spencer PS, Sproles DI, Swenberg JA, Suk WA, Sullivan RC, Tian R, Tennant RW, Todd SA, Tucker CJ, Van Houten B, Weis BK, Xuan S, Zarbl H; Members of the Toxicogenomics Research Consortium. (2005). "Standardizing global gene expression analysis between laboratories and across platforms". *Nat Methods* **2**: 351-356. PMID 15846362.
9. ^ Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. (1994). "Light-generated oligonucleotide arrays for rapid DNA sequence analysis.". *PNAS* **91**: 5022-5026. PMID 8197176.
10. ^ Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg JP, Ballin J, McCormick M, Norton J, Pollock T, Sumwalt T, Butcher L, Porter D, Molla M, Hall C, Blattner F, Sussman MR, Wallace RL, Cerrina F, Green RD. (2002). "Gene expression analysis using oligonucleotide arrays produced by maskless photolithography.". *Genome Res* **12**: 1749-1755. PMID 12421762.

11. ^ Shalon D, Smith SJ, Brown PO (1996). "A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization". *Genome Res* **6**: 639-645. PMID 8796352.
12. ^ Tang T, François N, Glatigny A, Agier N, Mucchielli MH, Aggerbeck L, Delacroix H (2007). "Expression ratio evaluation in two-colour microarray experiments is significantly improved by correcting image misalignment". *Bioinformatics* **23**: 2686-2691. PMID 17698492.
13. ^ http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/
14. ^ Wei C, Li J, Bumgarner RE. (2004). "Sample size for detectng differentially expressed genes in microarray experiments". *BMC Genomics* **5**: 87. PMID 15533245.

Protein microarray

From Wikipedia, the free encyclopedia

• *Find out more about navigating Wikipedia and finding information* •

A **protein microarray** is a piece of glass on which different molecules of protein have been affixed at separate locations in an ordered manner thus forming a microscopic array. These are used to identify protein-protein interactions, to identify the substrates of protein kinases, or to identify the targets of biologically active small molecules. The most common protein microarray is the antibody microarray, where antibodies are spotted onto the protein chip and are used as *capture molecules* to detect proteins from cell lysate solutions.

Related microarray technologies also include DNA microarrays, Antibody microarrays, Tissue microarrays and Chemical Compound Microarrays.

Contents

- 1 Applications
- 2 Types of chips
- 3 Production of protein arrays
  - 3.1 Artifacts to Avoid
- 4 Types of capture molecules
- 5 Detection methods
- 6 See also
- 7 References

- 8 External links

Applications

Protein microarrays (also **biochip**, **proteinchip**) are measurement devices used in biomedical applications to determine the presence and/or amount (referred to as quantitation) of proteins in biological samples, e.g. blood. They have the potential to be an important tool for proteomics research. Usually a multitude of different capture agents, most frequently monoclonal antibodies, are deposited on a chip surface (glass or silicon) in a miniature array. This format is often also referred to as a **microarray** (a more general term for chip based biological measurement devices).

Types of chips

There are several types of protein chips, the most common being glass slide chips and nano-well arrays.

Production of protein arrays

The proteins can be externally synthesised, purified and attached to the array. Alternatively they can be synthesised in-situ and directly attached to the array.

The proteins can be synthesised through biosynthesis, cell-free DNA expression or chemical synthesis. In-situ synthesis is possible with the latter two. With cell-free DNA expression, proteins are attached to the support right after their production. Peptides chemically procured by solid phase peptide synthesis are already attached to the support. Selective deprotection is carried out through lithographic methods or by the so-called SPOT-synthesis.

Artifacts to Avoid

1) To avoid results variability make sure you've got a very efficient lysis of the buffer.Use consistent sample processing conditions; 2) Many antibodies don't work well as capture reagents, even if they do work well in Western blotting and other denauring conditions. Some antibodies often bind poorly to intact proteins in a cell extract; 3) Different proteins like different solution conditions, so if you do not see binding it doesn't mean that there is no binding between the two partners in physiological conditions; 4) Adjust the solute conditions to avoid non-specific association: change salt concentration, pH, add 1% alignate; 5) on the array's surface the conjugated protein should be in the right conformation (i.e., folded, NOT denatured), anchored by the same amino acid (in the same orientation), and be kept away from the surface by a linker to avoid steric hinderance.

Types of capture molecules

Capture molecules used are most commonly antibodies; however, more recently there has been a push towards other types of capture molecules which are more similar in their nature such as peptides or aptamers. Antibodies have several problems including the fact that there are not antibodies for most proteins and also problems with specificity in some commercial antibody preparations. Nevertheless, antibodies still represent the most well-characterized and effective protein capture agent for microarrays. Recently, nucleic acids, receptors, enzymes, and proteins have been spotted onto chips and used as capture molecules. This allows a vast variety of experiments to be conducted on protein-protein interactions, and all other protein binding substrates.

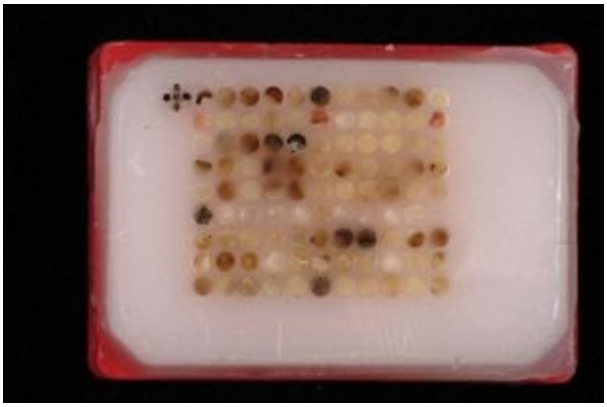
Detection methods

Although protein microarrays may use similar detection methods as DNA Microarrays, a problem is that protein concentrations in a biological sample may be many orders of magnitude different from that for mRNAs. Therefore, protein chip detection methods must have a much larger range of detection.

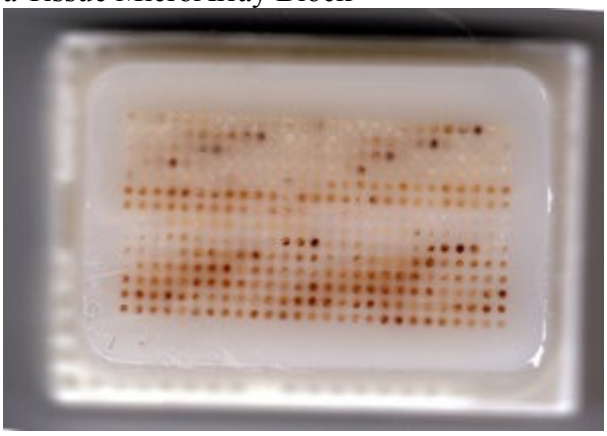
The preferred method of detection currently is fluorescence detection. Fluorescent detection is safe, sensitive, and can have a high resolution. The fluorescent detection method is compatible with standard microarray scanners, however some minor alterations to software may need to be made.

Tissue microarray

From Wikipedia, the free encyclopedia



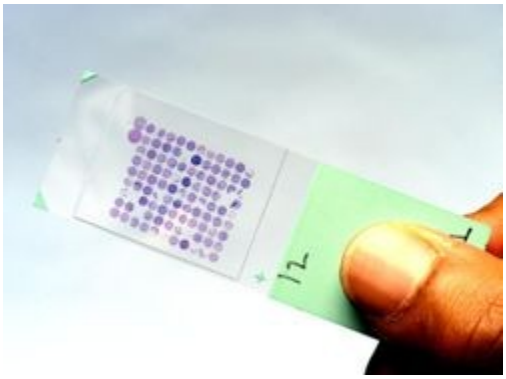
a Tissue MicroArray Block



0.6um core Tissue MicroArray Block

Tissue microarrays (also **TMA**s) consist of paraffin blocks in which up to 1000 separate tissue cores are assembled in array fashion to allow simultaneous histological analysis. The major limitations in molecular clinical analysis of tissues include the cumbersome nature of procedures, limited availability of diagnostic reagents and limited patient sample size. The technique of tissue microarray was developed to address these issues.

Multi-tissue blocks were first introduced by H. Battifora in 1986 with his so called “multitumor (sausage) tissue block" and modified in 1990 with its improvement, "the checkerboard tissue block" . In 1998 J. Kononen and collaborators developed the current technique, which uses a novel sampling approach to produce tissues of regular size and shape that can be more densely and precisely arrayed.



a Tissue MicroArray Section

In the tissue microarray technique, a hollow needle is used to remove tissue cores as small as 0.6 mm in diameter from regions of interest in paraffin embedded tissues such as clinical biopsies or tumor samples.

These tissue cores are then inserted in a recipient paraffin block in a precisely spaced, array pattern. Sections from this block are cut using a microtome, mounted on a microscope slide and then analyzed by any method of standard histological analysis. Each microarray block can be cut into 100 – 500 sections, which can be subjected to independent tests. Tests commonly employed in tissue microarray include immunohistochemistry, and fluorescent in situ hybridization. Tissue microarrays are particularly useful in analysis of cancer samples.

References

- Battifora H: The multitumor (sausage) tissue block: novel method for immunohistochemical antibody testing. Lab Invest 1986, 55:244-248.
- Battifora H, Mehta P: The checkerboard tissue block. An improved multitissue control block. Lab Invest 1990, 63:722-724.
- Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 1998, 4:844-847.

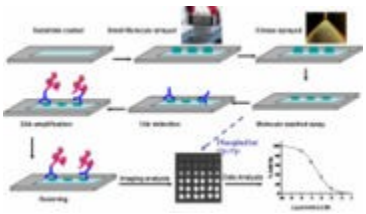
Chemical Compound Microarray

From Wikipedia, the free encyclopedia

• *Have questions? Find out how to ask questions and get answers.* •

A **Chemical compound microarray** is a collection of organic chemical compounds spotted on a solid surface, such as glass and plastic. This microarray format is very similar to DNA microarray, Protein microarray and Antibody microarray. In chemical genetics research, they are routinely used for searching proteins that binds with specific chemical compounds, and in general drug discovery research, they are used for searching potential drugs for therapeutic targets.

There are 3 different forms of chemical compound microarrays based on the fabrication method. The first form is to covalently immobilize the organic compounds on the solid surface with diverse linking techniques; this platform is usually called Small Molecule Microarray, which is invented and advanced by Dr. Stuart Schreiber and colleagues, [1]. The second form is to spot and dry organic compounds on the solid surface without immobilization, this platform has a commercial name as Micro Arrayed Compound Screening (μARCS), which is developed by scientists in Abbott Laboratories [2]. The last form is to spot organic compounds in a homogenous solution without immobilization and drying effect, this platform is developed by Dr. Scott Diamond [3] and later commercialized as DiscoveryDot™ technology by Reaction Biology Corporation [4].



DiscoveryDot Microarray for Kinase Profiling

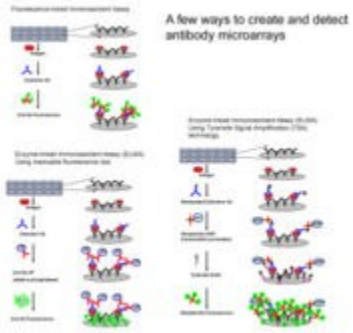
References

- Uttamchandani, M. *et al.* (2005) Small molecule microarrays, recent advances and applications. Curr Opin Chem Biol. 9, 4-13 [5].
- Walsh, D.P. and Chang, Y.T. (2004) Recent Advances in Small Molecule Microarrays, Applications and Technology. Comb Chem High Throughput Screen. 7, 557-564 [6].
- Hoever, M. and Zbinden, P. (2004) The evolution of microarrayed compound screening. Drug Discov. Today 9, 358-365.
- Gosalia, DN and Diamond, SL. (2003) Priting Chemical libraries on microarrays for fluid phase nanoliter reactions. Proc. Natl. Acad. Sci. USA, 100, 8721-8726 [7].
- Ma, H. *et al.* (2005) Nanoliter Homogenous Ultra High Throughput Screening Microarray for Lead Discoveries and IC50 Profiling. Assay Drug Dev. Technol. 3, 177-187 [8].
- Horiuchi, K.Y. *et al.* (2005) Microarrays for the functional analysis of the chemical-kinase interactome, accepted, J Biomol Screen. 11, 48-56 [9].
- Ma, H. and Horiuchi, K.Y. (2006) Chemical Microarray: a new tool for drug screening and discovery, Drug Discovery Today, 11, 661-668 [10].

Antibody microarray

From Wikipedia, the free encyclopedia

• *Learn more about using Wikipedia for research* •



Samples of antibody microarray creations and detections.

An **antibody microarray** is a specific form of protein microarrays, a collection of capture antibodies are spotted and fixed on a solid surface, such as glass, plastic and silicon chip for the purpose of detecting antigens. Antibody microarray is often used for detecting protein expressions from cell lysates in general research and special biomarkers from serum or urine for diagnostic applications.

Related microarray technologies also include Protein microarrays, DNA microarrays, Tissue microarrays and Chemical Compound Microarrays.

Gene chip analysis

From Wikipedia, the free encyclopedia

• *Learn more about using Wikipedia for research* •

Contents

- 1 Introduction
- 2 Data analysis
  - 2.1 Replicates

    - 2.2 Normalization
- 3 Quality control
    - 3.1 Filtering of flag
    - 3.2 Filtering of noise replicate
    - 3.3 Filtering of non significant gene
    - 3.4 Statistical analysis
    - 3.5 Clustering
    - 3.6 Hierarchical clustering
    - 3.7 K-mean clustering
- 4 Gene ontology study
- 5 Pathway analysis

- 6 Author

Introduction

Microarray is a powerful tool for genome analysis.It gives the global view of the genome analysis in a single experiment. Data analysis in the Microarray is a vital part as this part influences the final result. Each microarray experiment yields at least thousand data points. Each microarray study comprises multiple microarray experiments, each microarray study would give tens of thousands of data points. Since the volume of data growing exponential, the analysis becomes a challenging task. In general more the volume of data, more chances for erroneous results. Handling of such large volume of data requires high end computational infrastructure and programs that can handle multiple data formats. Already there are programs available for microarray data analysis on various platforms. But due to rapid development, diversity in microarray technology and different data formats, there is always need for comprehensive and complete microarray data analysis

Data analysis

Data analysis is the critical part of the whole analysis, since any error introduced in the data analysis part will result in biological insignificant results. In data analysis, the information from the raw data file further processed to get the meaningful biological results. This part includes data normalization; Flagging of the data, Averaging the ratio for replicates, Clustering of similarly expressed genes, etc. Each replicate data has to undergo normalization before further analysis. Normalization removes the non-biological variation between the samples. After the normalization, the ratio is calculated for each gene in the replicate. Based on the ratio, differentially regulated genes are determined. There are various statistical analysis also done for confidence analysis. Each replicate data also is examined for various experimental artifacts, bias by computing parameters related to intensity, background, flags, spot details, etc.

Replicates

It is important to note the necessity in conducting Microarray experiments in replicates. Like any other quantitative measurements, repeated experiments provide the ability to conduct confidence analysis and identify differentially expressed genes at a given level of confidence. More replicates provide more confidence in determining differentially expressed genes. In practice, three to five replicates would be an ideal.

Normalization

Normalization is required to standardize data and focus on biologically relevant changes. There are many sources of systematic variation in Microarray experiments that affect the measured gene expression levels such as Dye bias, Heat and light sensitivity, Efficiency of dye incorporation, Difference in the labeled cDNA Hybridization conditions, Scanning conditions, and Unequal quantities of starting RNA etc. Normalization is important step to Adjust data set for technical variation and removing relative abundance of gene expression profiles, this is only point where 1 and 2 color data analysis vary .The normalization method depends on the data. The basic idea behind all the normalization methods is that the expected mean intensity ratio between the two channels is one. If the observed mean intensity ratio deviates from one, the data is mathematically processed in such a way that the final observed mean intensity ratio becomes one. When the mean intensity ratio is adjusted to one, the distribution of the gene expression is centered so that genuine differentials can be identified

Quality control

Before doing analysis the biological variation must perform QC steps to determine if the data is fit for statistical test. Statistical tests are very sensitive to the nature of the input data.

Filtering of flag

Filtering on bad intensity spot is an important process of quality control For example; there is a certain limit of the scanner below which the intensity values cannot be trusted anymore. Typically, the lowest intensity value of the reliable data is about 100–200 for Affymetrix data and 100–1000 for cDNA Microarray data. These cut-offs are likely to change, as the scanners get more precise. The values below the cut-off point are usually removed (filtered) from the data, because they are likely to be artifacts.

Filtering of noise replicate

Filtering the noise replicate is one of the crucial parts in quality control. Experimental replicate should behave in similarly pattern. The replicates with noise should be eliminated before analysis .the noise replicate can be removed ANOVA statistical method

Filtering of non significant gene

Filtering of non significant is done to reduce the number of genes so that analysis could be done on selected genes. Nonsignificant genes were removed by specifying relative fold changewith respect to normal control. For over expressed and underexpressed genes values were given 2 & -2. As a result of the filtration few genes where retained. the remaining gene are then subjected to statistical analysis.

Statistical analysis

Statistical analysis plays a vital role in identifying the gene which is statistically significant expressed.

Clustering

Clustering is a data mining technique used to group the genes, which as similar expression patterns. Hierarchical clustering, k-mean clustering are widely used technique in microarray analysis.

Hierarchical clustering

Hierarchical clustering is a statistical method for finding relatively homogeneous Clusters. Hierarchical clustering consists of two separate phases. Initially, a distance matrix containing all the pair wise distances between the genes is calculated. Pearson's correlation or Spearman's correlation are often used as dissimilarity estimates, but other methods, like Manhattan distance or Euclidian distance can also be applied. If the genes on a single chip need to be clustered, the Euclidian distance is the correct choice, since at least two chips are needed for calculation of any correlation measures.After calculation of the initial distance matrix, the hierarchical clustering algorithm Either iteratively joins the two closest clusters starting from single clusters (Agglomerative, bottom-up approach) or iteratively partitions clusters starting from the complete set (divisive, top-down approach). After each step, a new distance matrix between the newly formed clusters and the other clusters is recalculated. If there are N cases, Hierarchical cluster analysis including:

• Single linkage (minimum method, nearest neighbor) • Complete linkage (maximum method, furthest neighbor) • Average Linkage (UPGMA).

K-mean clustering

K-mean clustering is an algorithm to classify or to group genes based on pattern into K number of group. K is positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus the purpose of K-mean clustering is to classify the data based on similar expression. (www.biostat.ucsf.edu).

Gene ontology study

Ontology study gives the biologically meaning full information like cellular location, molecular function and biological function about the gene which are differentially regulated in disease or drug treatment condition with respect to normal contol.

Pathway analysis

Pathway analysis gives the specific information about the pathway being affected in disease condition with reference to normal control. This pathway analysis also allows to identify the gene network and the genes how it regulated.

Cellular Microarray

From Wikipedia, the free encyclopedia

• Interested in contributing to Wikipedia? •

Cellular microarrays, are biological chips that allow interrogation of living cells on the surface of a solid support. The chips, spotted with varying materials, such as antibodies, proteins, or lipids, can interact with the cells, leading to their capture on specific spots. Combinations of different materials can be spotted in a given area, allowing not only cellular capture, when a specific interaction exists, but also the triggering of a cellular

response, change in phenotype, or detection of a response from the cell, such as a specific secreted factor. An example of cellular microarrays are PMHC Cellular Microarrays.

Cellular microarrays were developed by Daniel Chen, Yoav Soen, Dan Kraft, Patrick Brown and Mark Davis at Stanford University Medical Center.

PMHC Cellular Microarrays

From Wikipedia, the free encyclopedia

• *Have questions? Find out how to ask questions and get answers.* •

PMHC Cellular Microarrays are a type of Cellular Microarray that has been spotted with pMHC complexes peptide-MHC class I or peptide-MHC class II.

These biochips can be used to interrogate immune cells, particularly antigen-specific T cells, from clinical samples for what they are capable of recognizing. They can also be co-spotted with other molecules, such as antibodies that capture cytokines, allowing for high-throughput functional analysis of the captured T cells. Molecules spotted on a pMHC cellular microarray can be classified as capture molecules, detector molecules and effector molecules.

They were developed by Daniel Chen, Yoav Soen, Dan Kraft, Pat Brown and Mark Davis at Stanford University Medical Center.