Biological Techniques and Tools

Norbert Mercer

First Edition, 2012

ISBN 978-81-323-3131-5

Published by:
Research World
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1 - Cloning

Chapter 2 - Electrophoresis

Chapter 3 - In Situ Hybridization and Gene Knockout

Chapter 4 - Microscope

Chapter 5 - Northern Blot

Chapter 6 - Polymerase Chain Reaction

Chapter 7 - Western Blot and Southern Blot

Chapter 8 - X-ray Crystallography

Chapter- 1

Cloning

The sea anemone, *Anthopleura elegantissima* in process of cloning

Cloning in biology is the process of producing similar populations of genetically identical individuals that occurs in nature when organisms such as bacteria, insects or plants reproduce asexually. Cloning in biotechnology refers to processes used to create copies of DNA fragments (molecular cloning), cells (cell cloning), or organisms. The term also refers to the production of multiple copies of a product such as digital media or software.

The term *clone* is derived from *κλῶνος*, the Greek word for "trunk, branch", referring to the process whereby a new plant can be created from a twig. In horticulture, the spelling *clon* was used until the twentieth century; the final *e* came into use to indicate the vowel is a "long o" instead of a "short o". Since the term entered the popular lexicon in a more general context, the spelling *clone* has been used exclusively.

Molecular cloning

Molecular cloning refers to the process of making multiple molecules. Cloning is commonly used to amplify DNA fragments containing whole genes, but it can also be used to amplify any DNA sequence such as promoters, non-coding sequences and randomly fragmented DNA. It is used in a wide array of biological experiments and practical applications ranging from genetic fingerprinting to large scale protein production. Occasionally, the term cloning is misleadingly used to refer to the identification of the chromosomal location of a gene associated with a particular phenotype of interest, such as in positional cloning. In practice, localization of the gene to a chromosome or genomic region does not necessarily enable one to isolate or amplify the relevant genomic sequence. To amplify any DNA sequence in a living organism, that sequence must be linked to an origin of replication, which is a sequence of DNA capable of directing the propagation of itself and any linked sequence. However, a number of other features are needed and a variety of specialised cloning vectors (small piece of DNA into which a foreign DNA fragment can be inserted) exist that allow protein expression, tagging, single stranded RNA and DNA production and a host of other manipulations.

Cloning of any DNA fragment essentially involves four steps

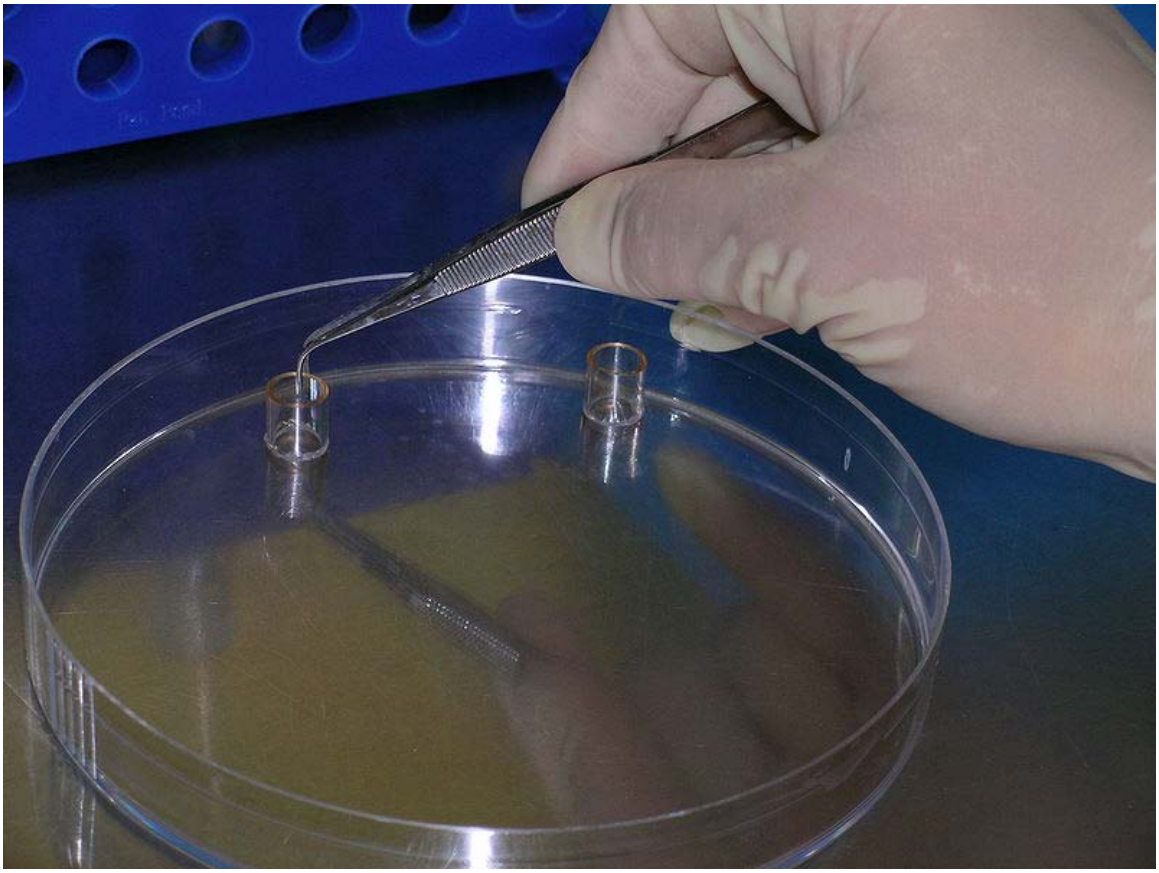
1. fragmentation - breaking apart a strand of DNA
2. ligation - gluing together pieces of DNA in a desired sequence
3. transfection - inserting the newly formed pieces of DNA into cells
4. screening/selection - selecting out the cells that were successfully transfected with the new DNA

Although these steps are invariable among cloning procedures a number of alternative routes can be selected, these are summarized as a 'cloning strategy'.

Initially, the DNA of interest needs to be isolated to provide a DNA segment of suitable size. Subsequently, a ligation procedure is used where the amplified fragment is inserted into a vector (piece of DNA). The vector (which is frequently circular) is linearised using restriction enzymes, and incubated with the fragment of interest under appropriate conditions with an enzyme called DNA ligase. Following ligation the vector with the insert of interest is transfected into cells. A number of alternative techniques are available, such as chemical sensitivation of cells, electroporation, optical injection and biolistics. Finally, the transfected cells are cultured. As the aforementioned procedures are of particularly low efficiency, there is a need to identify the cells that have been successfully transfected with the vector construct containing the desired insertion sequence in the required orientation. Modern cloning vectors include selectable antibiotic resistance markers, which allow only cells in which the vector has been transfected, to grow. Additionally, the cloning vectors may contain colour selection markers, which provide blue/white screening (??-factor complementation) on X-gal medium. Nevertheless, these selection steps do not absolutely guarantee that the DNA insert is present in the cells obtained. Further investigation of the resulting colonies must be

required to confirm that cloning was successful. This may be accomplished by means of PCR, restriction fragment analysis and/or DNA sequencing.

Unicellular organisms



Cloning cell-line colonies using cloning rings

Cloning a cell means to derive a population of cells from a single cell. In the case of unicellular organisms such as bacteria and yeast, this process is remarkably simple and essentially only requires the inoculation of the appropriate medium. However, in the case of cell cultures from multi-cellular organisms, cell cloning is an arduous task as these cells will not readily grow in standard media.

A useful tissue culture technique used to clone distinct lineages of cell lines involves the use of cloning rings (cylinders). According to this technique, a single-cell suspension of cells that have been exposed to a mutagenic agent or drug used to drive selection is plated at high dilution to create isolated colonies; each arising from a single and potentially clonal distinct cell. At an early growth stage when colonies consist of only a few of cells, sterile polystyrene rings (cloning rings), which have been dipped in grease are placed over an individual colony and a small amount of trypsin is added. Cloned cells are collected from inside the ring and transferred to a new vessel for further growth.

Cloning in stem cell research

Somatic cell nuclear transfer, known as SCNT, can also be used to create embryos for research or therapeutic purposes. The most likely purpose for this is to produce embryos for use in stem cell research. This process is also called "research cloning" or "therapeutic cloning." The goal is not to create cloned human beings (called "reproductive cloning"), but rather to harvest stem cells that can be used to study human development and to potentially treat disease. While a clonal human blastocyst has been created, stem cell lines are yet to be isolated from a clonal source.

Organism cloning

Organism cloning (also called reproductive cloning) refers to the procedure of creating a new multicellular organism, genetically identical to another. In essence this form of cloning is an asexual method of reproduction, where fertilization or inter-gamete contact does not take place. Asexual reproduction is a naturally occurring phenomenon in many species, including most plants and some insects. Scientists have made some major achievements with cloning, including the asexual reproduction of sheep and cows. There is a lot of ethical debate over whether or not cloning should be used. However, cloning, or asexual propagation, has been common practice in the horticultural world for hundreds of years.

Horticultural

The term *clone* is used in horticulture to mean all descendants of a single plant, produced by vegetative reproduction or apomixis. Many horticultural plant cultivars are clones, having been derived from a single individual, multiplied by some process other than sexual reproduction. As an example, some European cultivars of grapes represent clones that have been propagated for over two millennia. Other examples are potato and banana. Grafting can be regarded as cloning, since all the shoots and branches coming from the graft are genetically a clone of a single individual, but this particular kind of cloning has not come under ethical scrutiny and is generally treated as an entirely different kind of operation.

Many trees, shrubs, vines, ferns and other herbaceous perennials form clonal colonies. Parts of a large clonal colony often become detached from the parent, termed fragmentation, to form separate individuals. Some plants also form seeds is asexually, termed apomixis, e.g. dandelion.

Parthenogenesis

Clonal derivation exists in nature in some animal species and is referred to as parthenogenesis (reproduction of an organism by itself without a mate). This is an asexual form of reproduction that is only found in females of some insects, crustaceans and lizards. The growth and development occurs without fertilization by a male. In

plants, parthenogenesis means the development of an embryo from an unfertilized egg cell, and is a component process of apomixis. In species that use the XY sex-determination system, the offspring will always be female. An example is the "Little Fire Ant" (*Wasmannia auropunctata*), which is native to Central and South America but has spread throughout many tropical environments.

Artificial cloning of organisms

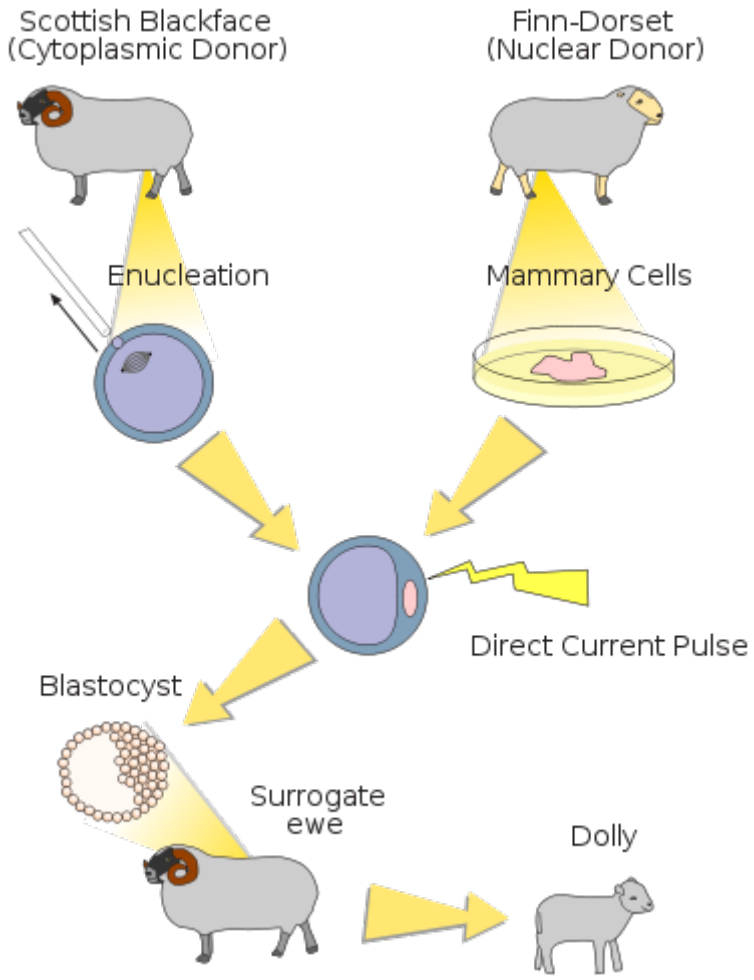
Artificial cloning of organisms may also be called *reproductive cloning*.

Methods

Reproductive cloning generally uses "somatic cell nuclear transfer" (SCNT) to create animals that are genetically identical. This process entails the transfer of a nucleus from a donor adult cell (somatic cell) to an egg that has no nucleus. If the egg begins to divide normally it is transferred into the uterus of the surrogate mother. Such clones are not strictly identical since the somatic cells may contain mutations in their nuclear DNA. Additionally, the mitochondria in the cytoplasm also contains DNA and during SCNT this DNA is wholly from the donor egg, thus the mitochondrial genome is not the same as that of the nucleus donor cell from which it was produced. This may have important implications for cross-species nuclear transfer in which nuclear-mitochondrial incompatibilities may lead to death.

Artificial *embryo splitting* or *embryo twinning* may also be used as a method of cloning, where an embryo is split in the maturation before embryo transfer. It is optimally performed at the 6- to 8-cell stage, where it can be used as an expansion of IVF to increase the number of available embryos. If both embryos are successful, it gives rise to monozygotic (identical) twins.

Dolly the Sheep



Dolly, a Finn-Dorset ewe, was the first mammal to have been successfully cloned from an adult cell. She was cloned at the Roslin Institute in Scotland and lived there from her birth in 1996 until her death in 2003 when she was six. Her stuffed remains were placed at Edinburgh's Royal Museum, part of the National Museums of Scotland.

Dolly was publicly significant because the effort showed that the genetic material from a specific adult cell, programmed to express only a distinct subset of its genes, can be reprogrammed to grow an entirely new organism. Before this demonstration, it had been shown by John Gurdon that nuclei from differentiated cells could give rise to an entire organism after transplantation into an enucleated egg. However, this concept was not yet demonstrated in a mamallian system.

Cloning Dolly the sheep had a low success rate per fertilized egg; she was born after 237 eggs were used to create 29 embryos, which only produced three lambs at birth, only one of which lived. Seventy calves have been created and one third of them died young; Prometea took 277 attempts. Notably, although the first clones were frogs, no adult cloned frog has yet been produced from a somatic adult nucleus donor cell.

There were early claims that Dolly the Sheep had pathologies resembling accelerated aging. Scientists speculated that Dolly's death in 2003 was related to the shortening of telomeres, DNA-protein complexes that protect the end of linear chromosomes. However, other researchers, including Ian Wilmut who led the team that successfully cloned Dolly, argue that Dolly's early death due to respiratory infection was unrelated to deficiencies with the cloning process.

Species cloned

The modern cloning techniques involving nuclear transfer have been successfully performed on several species. Landmark experiments in chronological order:

- Tadpole: (1952) Many scientists questioned whether cloning had actually occurred and unpublished experiments by other labs were not able to reproduce the reported results.
- Carp: (1963) In China, embryologist Tong Dizhou produced the world's first cloned fish by inserting the DNA from a cell of a male carp into an egg from a female carp. He published the findings in a Chinese science journal.
- Mice: (1986) A mouse was the first mammal successfully cloned from an early embryonic cell. Soviet scientists Chaylakhyan, Veprencev, Sviridova, and Nikitin had the mouse "Masha" cloned. Research was published in the magazine "Biofizika" volume XXXII, issue 5 of 1987.
- Sheep: (1996) From early embryonic cells by Steen Willadsen. Megan and Morag cloned from differentiated embryonic cells in June 1995 and Dolly the sheep from a somatic cell in 1997.
- Rhesus Monkey: Tetra (January 2000) from embryo splitting
- Gaur: (2001) was the first endangered species cloned.
- Cattle: Alpha and Beta (males, 2001) and (2005) Brazil
- Cat: CopyCat "CC" (female, late 2001), Little Nicky, 2004, was the first cat cloned for commercial reasons
- Dog: Snuppy, a male Afghan hound was the first cloned dog (2005).
- Rat: Ralph, the first cloned rat (2003)
- Mule: Idaho Gem, a john mule born 4 May 2003, was the first horse-family clone.
- Horse: Prometea, a Haflinger female born 28 May 2003, was the first horse clone.
- Water Buffalo: Samrupa was the first cloned water buffalo. It was born on February 6, 2009, at India's Karnal National Diary Research Institute but died five days later due to lung infection.
- Camel: (2009) Injaz, is the first cloned camel.

Human cloning

Human cloning is the creation of a genetically identical copy of an existing or previously existing human. The term is generally used to refer to *artificial* human cloning; human clones in the form of identical twins are commonplace, with their cloning occurring during the natural process of reproduction. There are two commonly discussed types of human cloning: therapeutic cloning and reproductive cloning. Therapeutic cloning involves cloning adult cells for use in medicine and is an active area of research. Reproductive cloning would involve making cloned humans. A third type of cloning called replacement cloning is a theoretical possibility, and would be a combination of therapeutic and reproductive cloning. Replacement cloning would entail the replacement of an extensively damaged, failed, or failing body through cloning followed by whole or partial brain transplant.

The various forms of human cloning are controversial. There have been numerous demands for all progress in the human cloning field to be halted. Most scientific, governmental and religious organizations oppose reproductive cloning. The American Association for the Advancement of Science (AAAS) and other scientific organizations have made public statements suggesting that human reproductive cloning be banned until safety issues are resolved. Serious ethical concerns have been raised by the future possibility of harvesting organs from clones. Some people have considered the idea of growing organs separately from a human organism - in doing this, a new organ supply could be established without the moral implications of harvesting them from humans. Research is also being done on the idea of growing organs that are biologically acceptable to the human body inside of other organisms, such as pigs or cows, then transplanting them to humans, a form of xenotransplantation.

The first hybrid human clone was created in November 1998, by American Cell Technologies. It was created from a man's leg cell, and a cow's egg whose DNA was removed. It was destroyed after 12 days. Since a normal embryo implants at 14 days, Dr Robert Lanza, ACT's director of tissue engineering, told the Daily Mail newspaper that the embryo could not be seen as a person before 14 days. While making an embryo, which may have resulted in a complete human had it been allowed to come to term, according to ACT: "[ACT's] aim was 'therapeutic cloning' not 'reproductive cloning'"

On January, 2008, Wood and Andrew French, Stemagen's chief scientific officer in California, announced that they successfully created the first 5 mature human embryos using DNA from adult skin cells, aiming to provide a source of viable embryonic stem cells. Dr. Samuel Wood and a colleague donated skin cells, and DNA from those cells was transferred to human eggs. It is not clear if the embryos produced would have been capable of further development, but Dr. Wood stated that if that were possible, using the technology for reproductive cloning would be both unethical and illegal. The 5 cloned embryos, created in Stemagen Corporation lab, in La Jolla, were destroyed.

Ethical issues of cloning

Because of recent technological advancements, the cloning of animals (and potentially humans) has been an issue. The Catholic Church and many religious organizations oppose all forms of cloning, on the grounds that life begins at conception. Judaism does not equate life with conception and, though some question the wisdom of cloning, Orthodox rabbis generally find no firm reason in Jewish law and ethics to object to cloning. From the standpoint of classical liberalism, concerns also exist regarding the protection of the identity of the individual and the right to protect one's genetic identity.

Gregory Stock is a scientist and outspoken critic against restrictions on cloning research. Bioethicist Gregory Pence also attacks the idea of criminalizing attempts to clone humans.

The social implications of an artificial human production scheme were famously explored in Aldous Huxley's novel *Brave New World.*

On December 28, 2006, the U.S. Food and Drug Administration (FDA) approved the consumption of meat and other products from cloned animals. Cloned-animal products were said to be virtually indistinguishable from the non-cloned animals. Furthermore, companies would not be required to provide labels informing the consumer that the meat comes from a cloned animal.

Critics have raised objections to the FDA's approval of cloned-animal products for human consumption, arguing that the FDA's research was inadequate, inappropriately limited, and of questionable scientific validity. Several consumer-advocate groups are working to encourage a tracking program that would allow consumers to become more aware of cloned-animal products within their food.

Joseph Mendelson, legal director of the Center for Food Safety, said that cloned food still should be labeled since safety and ethical issues about it remain questionable.

Carol Tucker Foreman, director of food policy at the Consumer Federation of America, stated that FDA does not consider the fact that the results of some studies revealed that cloned animals have increased rates of mortality and deformity at birth.

Cloning extinct and endangered species

Cloning, or more precisely, the reconstruction of functional DNA from extinct species has, for decades, been a dream of some scientists. The possible implications of this were dramatized in the best-selling novel by Michael Crichton and high budget Hollywood thriller *Jurassic Park*. In real life, one of the most anticipated targets for cloning was once the Woolly Mammoth, but attempts to extract DNA from frozen mammoths have been unsuccessful, though a joint Russo-Japanese team is currently working toward this goal.

In 2001, a cow named Bessie gave birth to a cloned Asian gaur, an endangered species, but the calf died after two days. In 2003, a banteng was successfully cloned, followed by three African wildcats from a thawed frozen embryo. These successes provided hope that similar techniques (using surrogate mothers of another species) might be used to clone extinct species. Anticipating this possibility, tissue samples from the last *bucardo* (Pyrenean Ibex) were frozen in liquid nitrogen immediately after it died in 2000. Researchers are also considering cloning endangered species such as the giant panda, ocelot, and cheetah. The "Frozen Zoo" at the San Diego Zoo now stores frozen tissue from the world's rarest and most endangered species.

In 2002, geneticists at the Australian Museum announced that they had replicated DNA of the Thylacine (Tasmanian Tiger), extinct about 65 years previous, using polymerase chain reaction. However, on February 15, 2005 the museum announced that it was stopping the project after tests showed the specimens' DNA had been too badly degraded by the (ethanol) preservative. On 15 May 2005 it was announced that the Thylacine project would be revived, with new participation from researchers in New South Wales and Victoria.

In January 2009, for the first time, an extinct animal, the Pyrenean ibex mentioned above was cloned, at the Centre of Food Technology and Research of Aragon, using the preserved DNA of the skin samples from 2001 and domestic goat egg-cells. (The ibex died shortly after birth due to physical defects in its lungs.) One of the continuing obstacles in the attempt to clone extinct species is the need for nearly perfect DNA. Cloning from a single specimen could not create a viable breeding population in sexually reproducing animals. Furthermore, even if males and females were to be cloned, the question would remain open whether they would be viable at all in the absence of parents that could teach or show them their natural behavior.

Cloning endangered species is a highly ideological issue. Many conservation biologists and environmentalists vehemently oppose cloning endangered species — mainly because they think it may deter donations to help preserve natural habitat and wild animal populations. The "rule-of-thumb" in animal conservation is that, if it is still feasible to conserve habitat and viable wild populations, breeding in captivity should not be undertaken in isolation.

In a 2006 review, David Ehrenfeld concluded that cloning in animal conservation is an experimental technology that, at its state in 2006, could not be expected to work except by pure chance and utterly failed a cost-benefit analysis. Furthermore, he said, it is likely to siphon funds from established and working projects and does not address any of the issues underlying animal extinction (such as habitat destruction, hunting or other overexploitation, and an impoverished gene pool). While cloning technologies are well-established and used on a regular basis in plant conservation, care must be taken to ensure genetic diversity. He concluded:

Vertebrate cloning poses little risk to the environment, but it can consume scarce conservation resources, and its chances of success in preserving species seem poor. To

date, the conservation benefits of transgenics and vertebrate cloning remain entirely theoretical, but many of the risks are known and documented. Conservation biologists should devote their research and energies to the established methods of conservation, none of which require transgenics or vertebrate cloning.

Chapter- 2

Electrophoresis

Electrophoresis is the motion of dispersed particles relative to a fluid under the influence of a spatially uniform electric field. This electrokinetic phenomenon was observed for the first time in 1807 by Reuss, who noticed that the application of a constant electric field caused clay particles dispersed in water to migrate. It is ultimately caused by the presence of a charged interface between the particle surface and the surrounding fluid.

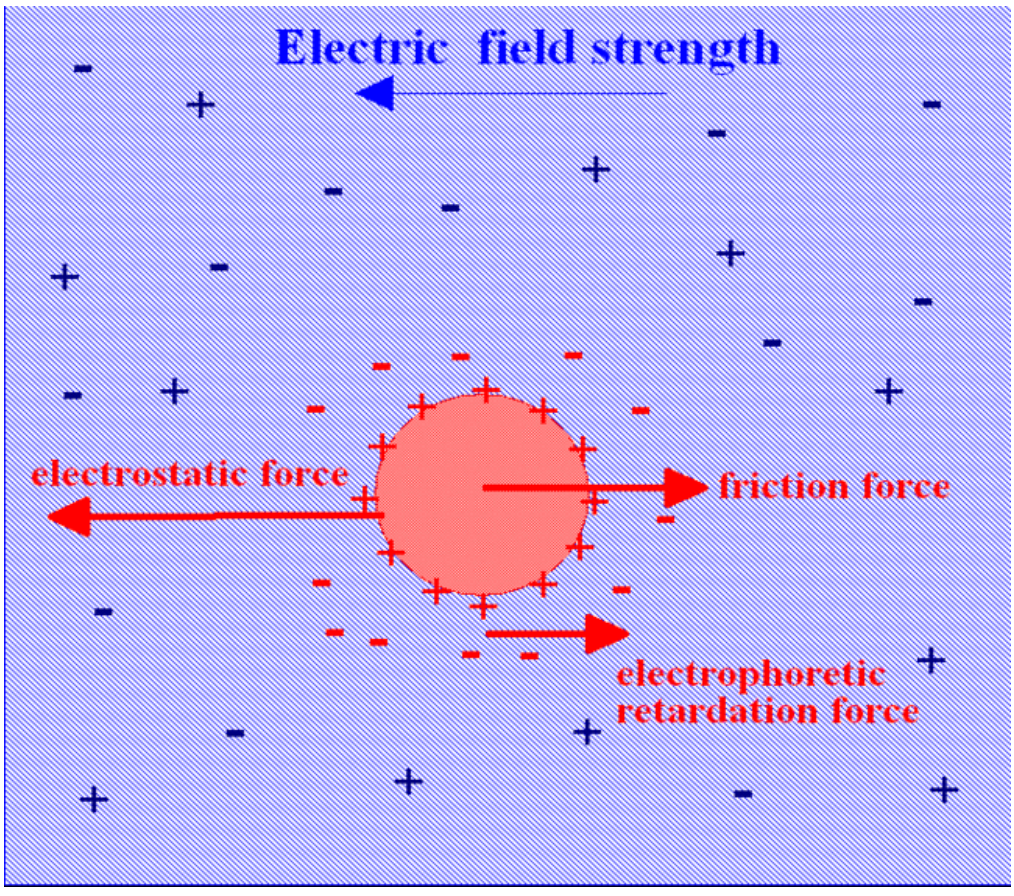


Illustration of electrophoresis

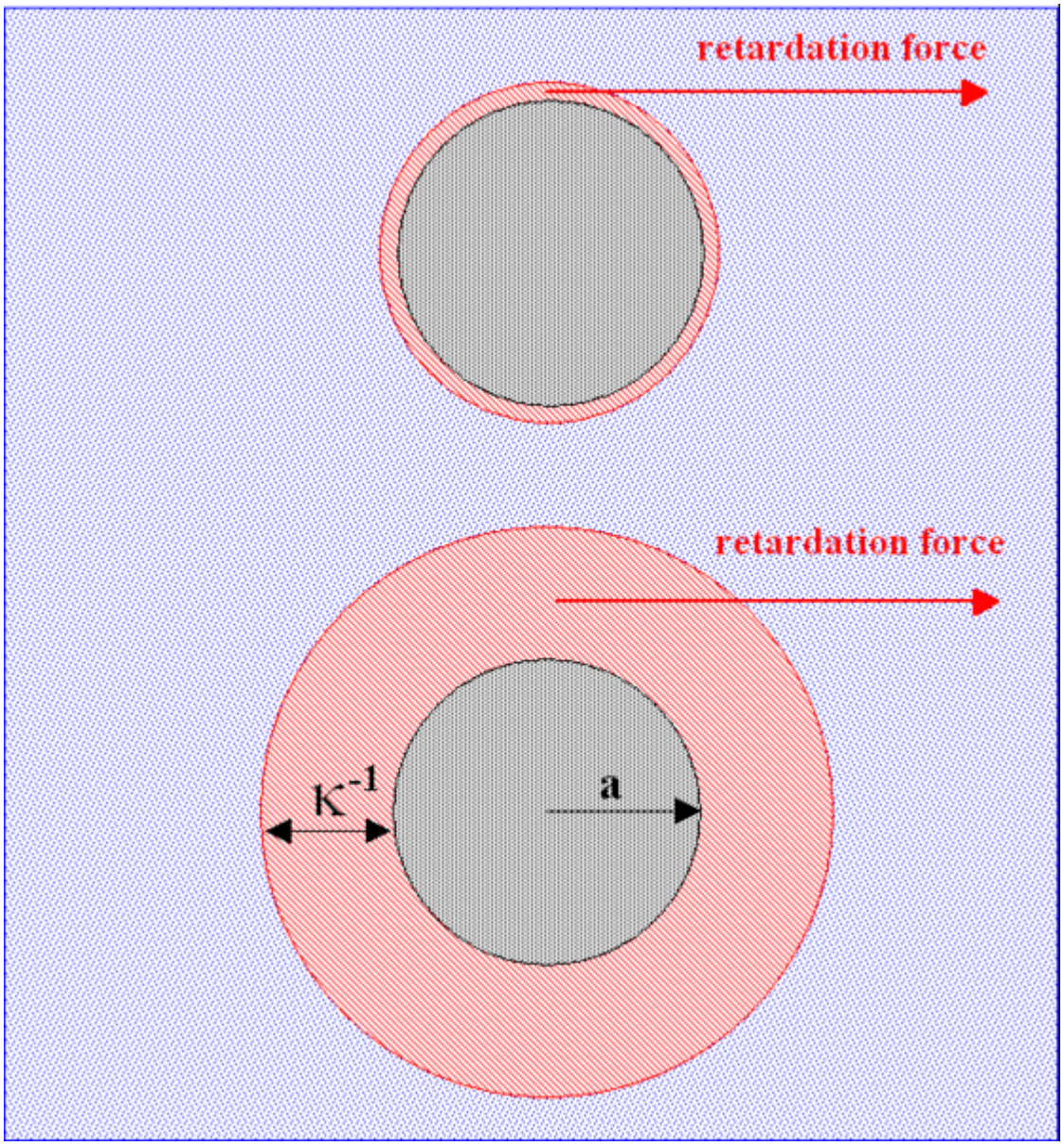


Illustration of electrophoresis retardation

Theory

The dispersed particles have an electric surface charge, on which an external electric field exerts an electrostatic Coulomb force. According to the double layer theory, all surface charges in fluids are screened by a diffuse layer of ions, which has the same absolute charge but opposite sign with respect to that of the surface charge. The electric field also exerts a force on the ions in the diffuse layer which has direction opposite to that acting on the surface charge. This latter force is not actually applied to the particle, but to the ions in the diffuse layer located at some distance from the particle surface, and part of it

is transferred all the way to the particle surface through viscous stress. This part of the force is also called electrophoretic retardation force.

This photograph shows a technician using an electrophoresis apparatus to separate proteins by molecular weight.

Considering the hydrodynamic friction on the moving particles due to the viscosity of the dispersant, in the case of low Reynolds number and moderate electric field strength *E*, the velocity of a dispersed particle *v* is simply proportional to the applied field, which leaves the electrophoretic mobility μ_e defined as:

$$\mu_e = \frac{v}{E}.$$

The most known and widely used theory of electrophoresis was developed in 1903 by Smoluchowski

$$\mu_e = \frac{\varepsilon_r \varepsilon_0 \zeta}{\eta},$$

where ε_r is the dielectric constant of the dispersion medium, ε_0 is the permittivity of free space ($C^2\ N^{-1}\ m^{-2}$), η is dynamic viscosity of the dispersion medium (Pa s), and ζ is zeta potential (i.e., the electrokinetic potential of the slipping plane in the double layer).

The Smoluchowski theory is very powerful because it works for dispersed particles of any shape at any concentration. Unfortunately, it has limitations on its validity. It follows, for instance, from the fact that it does not include Debye length κ^{-1}. However, Debye length must be important for electrophoresis, as follows immediately from the Figure on the right. Increasing thickness of the double layer (DL) leads to removing point of retardation force further from the particle surface. The thicker DL, the smaller retardation force must be.

Detailed theoretical analysis proved that the Smoluchowski theory is valid only for sufficiently thin DL, when particle radius *a* is much greater than the Debye length :

$$a\kappa \gg 1.$$

This model of "thin Double Layer" offers tremendous simplifications not only for electrophoresis theory but for many other electrokinetic theories. This model is valid for most aqueous systems because the Debye length is only a few nanometers there. It breaks only for nano-colloids in solution with ionic strength close to water

The Smoluchowski theory also neglects contribution of surface conductivity. This is expressed in modern theory as condition of small Dukhin number

$$Du << 1$$

In the effort of expanding the range of validity of electrophoretic theories, the opposite asymptotic case was considered, when Debye length is larger than particle radius:

$$\kappa a < 1.$$

Under this condition of a "thick Double Layer", Huckel predicted the following relation for electrophoretic mobility:

$$\mu_e = \frac{2\varepsilon_r \varepsilon_0 \zeta}{3\eta}.$$

This model can be useful for some nanoparticles and non-polar fluids, where Debye length is much larger than in the usual cases.

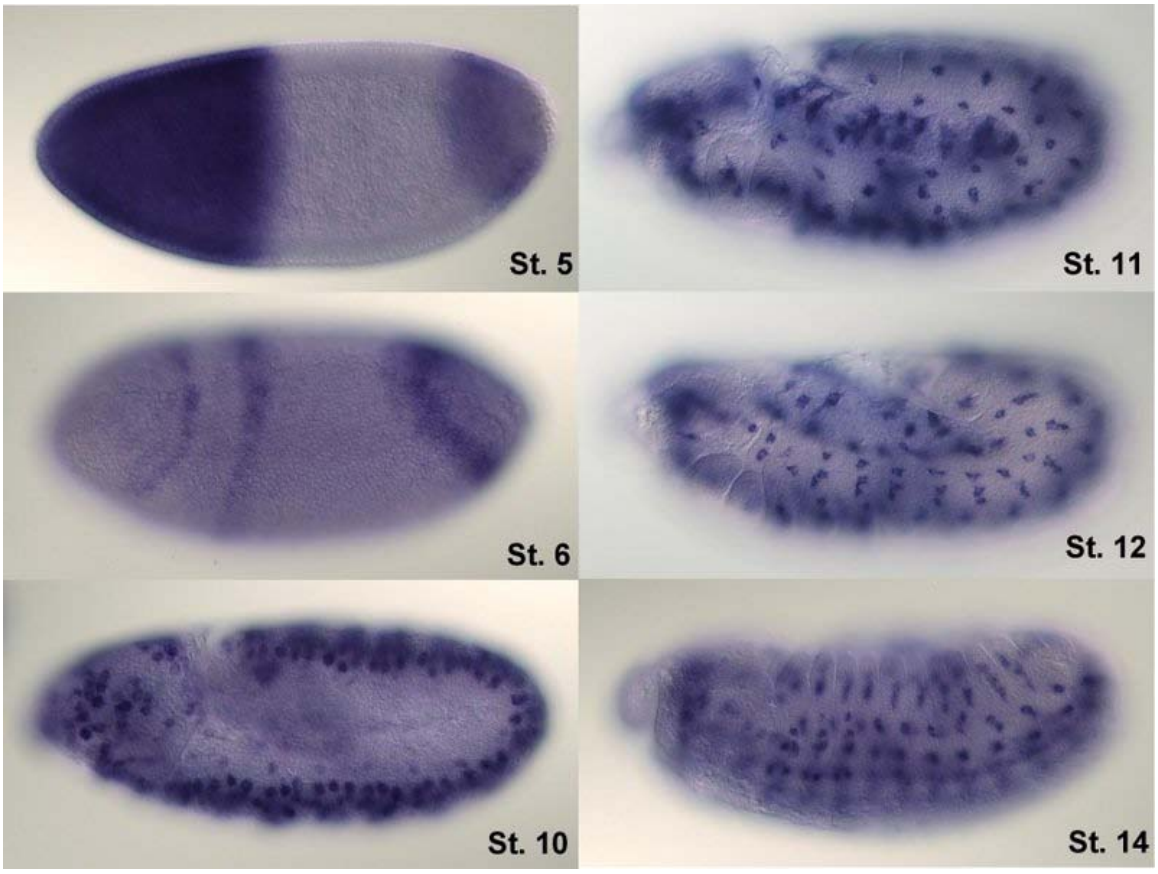
There are several analytical theories that incorporate surface conductivity and eliminate the restriction of a small Dukhin number, pioneered by Overbeek and Booth. Modern, rigorous theories valid for any Zeta potential and often any *κa* stem mostly from Dukhin-Semenikhin theory. In the **thin Double Layer** limit, these theories confirm the numerical solution to the problem provided by O'Brien and White.

Recent molecular dynamics required for electrophoresis to occur, and that even neutral particles can migrate in an electric field due to the molecular structure of water at the interface.

Chapter- 3

In Situ Hybridization and Gene Knockout

In situ hybridization



In situ hybridization of wild type *Drosophila* embryos at different developmental stages for the RNA from a gene called hunchback.

In situ hybridization (ISH) is a type of hybridization that uses a labeled complementary DNA or RNA strand (i.e., probe) to localize a specific DNA or RNA sequence in a

portion or section of tissue (*in situ*), or, if the tissue is small enough (e.g. plant seeds, *Drosophila* embryos), in the entire tissue (whole mount ISH). This is distinct from immunohistochemistry, which usually localizes proteins in tissue sections. DNA ISH can be used to determine the structure of chromosomes. Fluorescent DNA ISH (FISH) can, for example, be used in medical diagnostics to assess chromosomal integrity. RNA ISH (hybridization histochemistry) is used to measure and localize mRNAs and other transcripts within tissue sections or whole mounts.

Process

For hybridization histochemistry, sample cells and tissues are usually treated to fix the target transcripts in place and to increase access of the probe. As noted above, the probe is either a labeled complementary DNA or, now most commonly, a complementary RNA (riboprobe). The probe hybridizes to the target sequence at elevated temperature, and then the excess probe is washed away (after prior hydrolysis using RNase in the case of unhybridized, excess RNA probe). Solution parameters such as temperature, salt and/or detergent concentration can be manipulated to remove any non-identical interactions (i.e. only exact sequence matches will remain bound). Then, the probe that was labeled with either radio-, fluorescent- or antigen-labeled bases (e.g., digoxigenin) is localized and quantitated in the tissue using either autoradiography, fluorescence microscopy or immunohistochemistry, respectively. ISH can also use two or more probes, labeled with radioactivity or the other non-radioactive labels, to simultaneously detect two or more transcripts.

Basic Steps for Digoxigenin-labeled probes

1. permeabilisation of cells with proteinase K to open cell membranes (around 25 minutes, not needed for tissue sections or some early-stage embryos)
2. binding of mRNAs to marked RNA probe (usually overnight)
3. antibody-phosphatase binding to RNA-probe (some hours)
4. staining of antibody (e.g. with alkaline phosphatase)

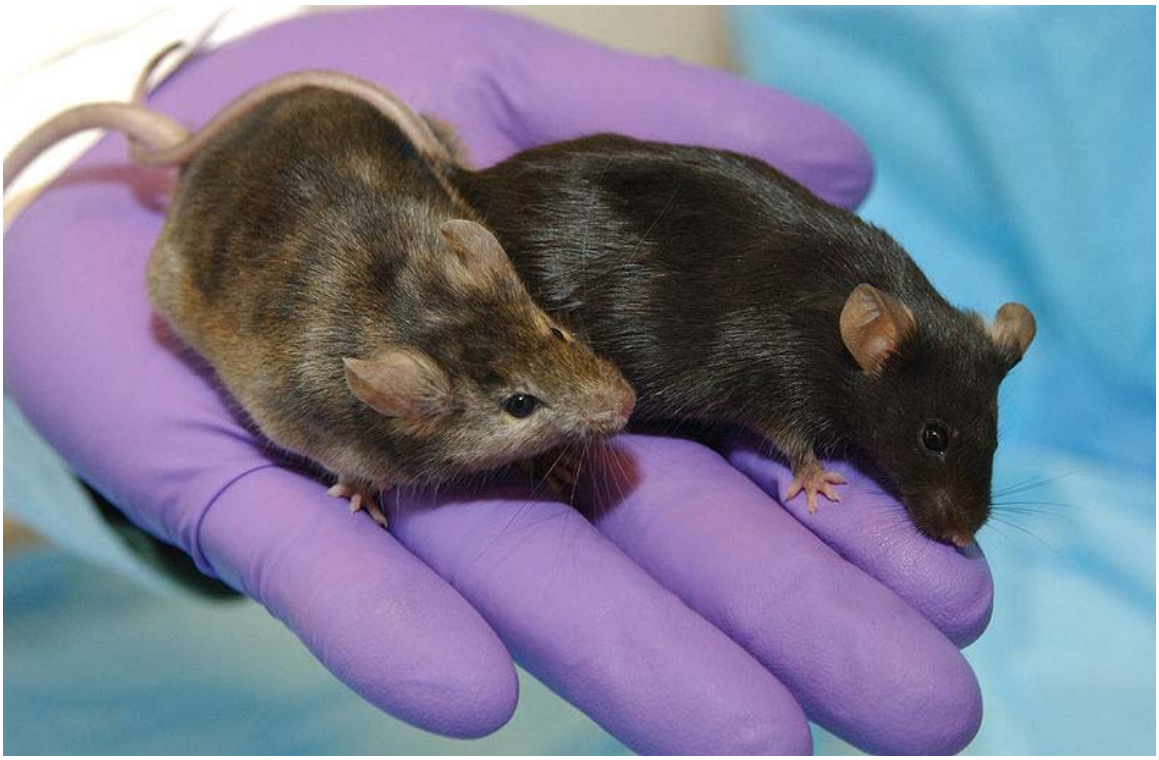
The protocol takes around 2-3 days and takes some time to set up. Some companies sell robots to automate the process. As a result, large-scale screenings have been conducted in laboratories on thousands of genes.

Gene knockout

A **gene knockout** (abbreviation: **KO**) is a genetic technique in which one of an organism's genes is made inoperative ("knocked out" of the organism). Also known as **knockout organisms** or simply **knockouts**, they are used in learning about a gene that has been sequenced, but which has an unknown or incompletely known function. Researchers draw inferences from the difference between the knockout organism and normal individuals.

The term also refers to the process of creating such an organism, as in "knocking out" a gene. The technique is essentially the opposite of a Gene Knock-in. Knockout is often abbreviated as **KO**. Knocking out two genes simultaneously in an organism is known as a **double knockout** (**DKO**). Similarly the terms **triple knockout** (**TKO**) and **quadruple knockouts** (**QKO**) are used to describe 3 or 4 knocked out genes, respectively.

Method

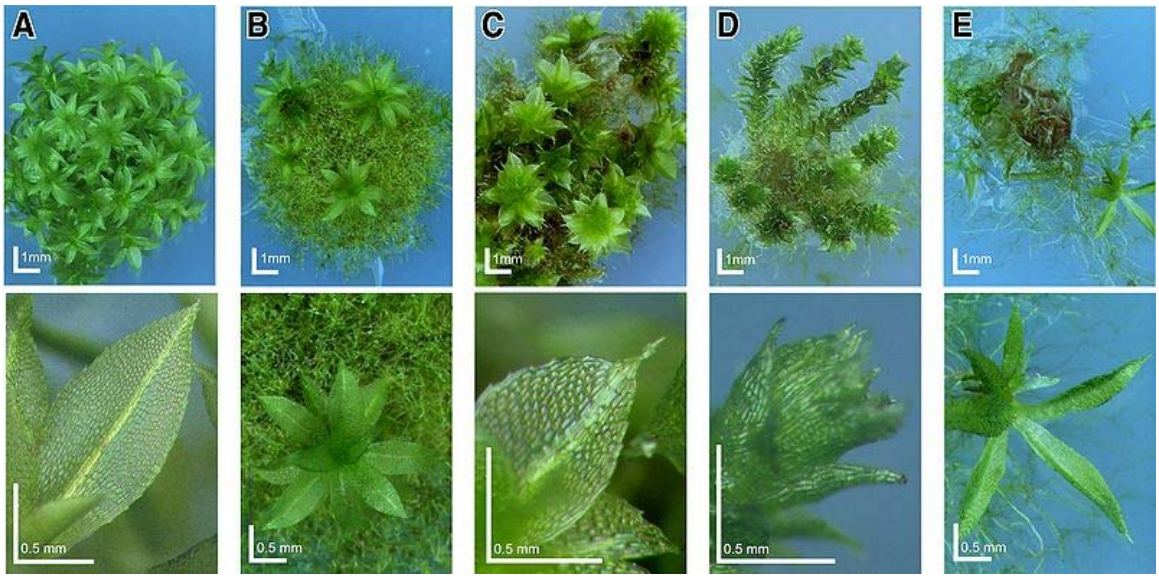


A laboratory mouse in which a gene affecting hair growth has been knocked out (left), is shown next to a normal lab mouse.

Knockout is accomplished through a combination of techniques, beginning in the test tube with a plasmid, a bacterial artificial chromosome or other DNA construct, and proceeding to cell culture. Individual cells are genetically transformed with the DNA construct. Often the goal is to create a transgenic animal that has the altered gene. If so,

embryonic stem cells are genetically transformed and inserted into early embryos. Resulting animals with the genetic change in their germline cells can then often pass the gene knockout to future generations.

To create knockout moss, transfection of protoplasts is the preferred method. Such transformed *Physcomitrella*-protoplasts directly regenerate into fertile moss plants. Already eight weeks after transfection the plants can be screened for gene targeting via PCR.



Wild-type Physcomitrella and knockout-mosses: Deviating phenotypes induced in gene-disruption library transformants. Physcomitrella wild-type and transformed plants were grown on minimal Knop medium to induce differentiation and development of gametophores. For each plant, an overview (upper row, scale bar corresponds to 1 mm) and a close-up (bottom row, scale bar equals 0.5 mm) is shown. A, Haploid wild-type moss plant completely covered with leafy gametophores and close-up of wild-type leaf. B-D, Different Mutants.

The construct is engineered to recombine with the target gene, which is accomplished by incorporating sequences from the gene itself into the construct. Recombination then occurs in the region of that sequence within the gene, resulting in the insertion of a foreign sequence to disrupt the gene. With its sequence interrupted, the altered gene in most cases will be translated into a nonfunctional protein, if it is translated at all.

A knockout mouse (left) that is a model of obesity, compared with a normal mouse.

A conditional knockout allows gene deletion in a tissue or time specific manner. This is done by introducing short sequences called loxP sites around the gene. These sequences will be introduced into the germ-line via the same mechanism as a knock-in. This germ-line can then be crossed to another germline containing Cre-recombinase which is a bacterial enzyme that can recognize these sequences, recombines them and deletes the gene flanked by these sites.

Because the desired type of DNA recombination is a rare event in the case of most cells and most constructs, the foreign sequence chosen for insertion usually includes a reporter. This enables easy selection of cells or individuals in which knockout was successful. Sometimes the DNA construct inserts into a chromosome without the desired homologous recombination with the target gene. To eliminate such cells, the DNA construct often contains a second region of DNA that allows such cells to be identified and discarded.

In diploid organisms, which contain two alleles for most genes, and may as well contain several related genes that collaborate in the same role, additional rounds of transformation and selection are performed until every targeted gene is knocked out. Selective breeding may be required to produce homozygous knockout animals.

Chapter- 4

Microscope

Microscope

Uses	Small sample observation
Notable experiments	Discovery of cells
Inventor	Hans Lippershey Zacharias Janssen
Related items	Optical microscope Electron microscope

18th century microscopes from the Musée des Arts et Métiers, Paris.

Microscope

A **microscope** (from the Greek: μικρός, *mikrós*, "small" and σκοπεῖν, *skopeîn*, "to look" or "see") is an instrument to see objects too small for the naked eye. The science of investigating small objects using such an instrument is called microscopy. Microscopic means invisible to the eye unless aided by a microscope.

There are many types of microscopes, the most common and first to be invented is the optical microscope which uses light to image the sample. Other major types of microscopes are the electron microscope (both the transmission electron microscope and the scanning electron microscope) and the various types of scanning probe microscope.

History

The first microscope to be developed was the optical microscope, although the original inventor is not easy to identify. An early microscope was made in 1590 in Middelburg, Netherlands. Two eyeglass makers are variously given credit: Hans Lippershey (who developed an early telescope) and Hans Janssen. Giovanni Faber coined the name **microscope** for Galileo Galilei's compound microscope in 1625 (Galileo had called it the "*occhiolino*" or "*little eye*").

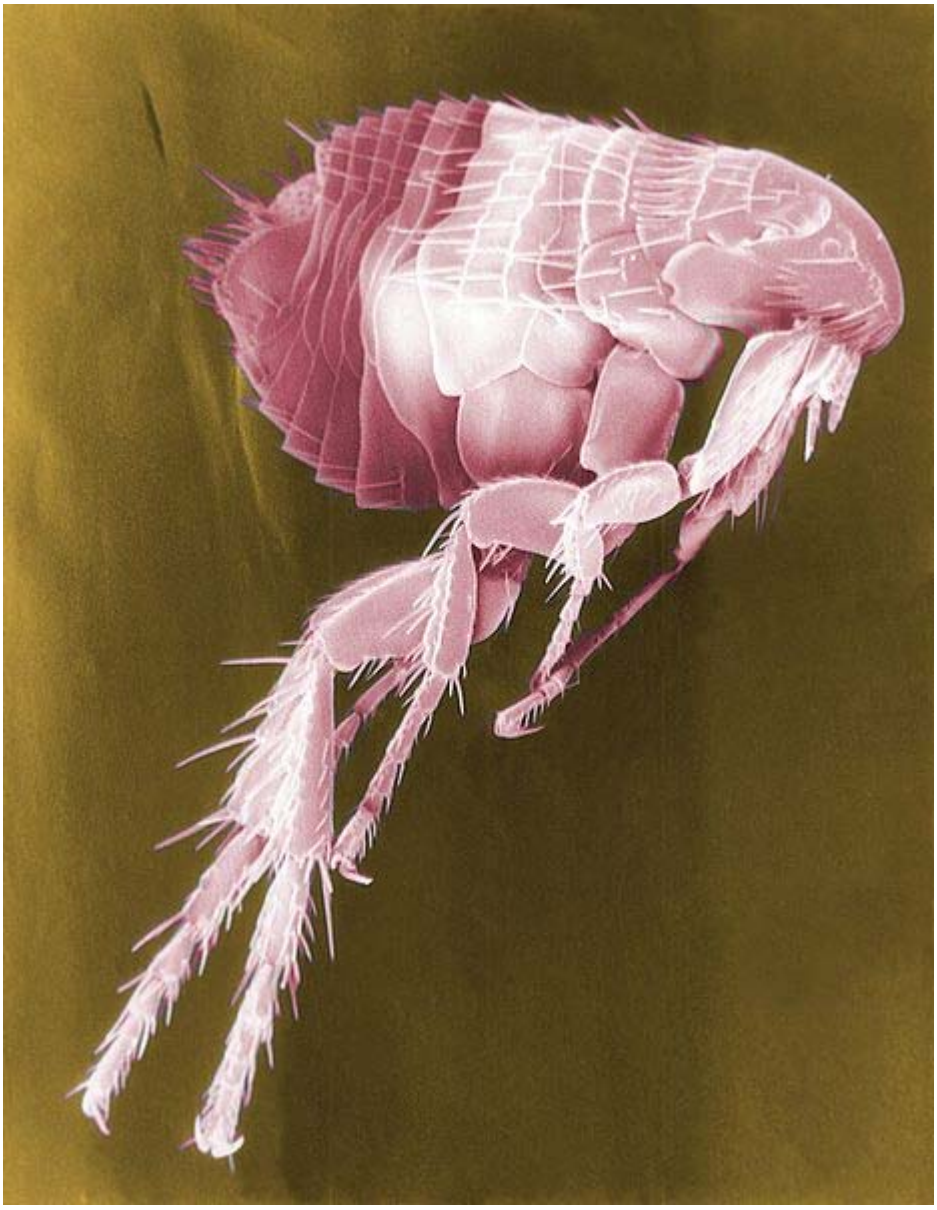
The rise of modern light microscopy

The first detailed account of the interior construction of living tissue based on the use of a microscope did not appear until 1644, in Giambattista Odierna's *L'occhio della mosca*, or *The Fly's Eye*.

It was not until the 1660s and 1670s that the microscope was used extensively for research in Italy, Holland and England. Marcelo Malpighi in Italy began the analysis of biological structures beginning with the lungs. Robert Hooke's *Micrographia* had a huge impact, largely because of its impressive illustrations. The greatest contribution came from Antoni van Leeuwenhoek who discovered red blood cells and spermatozoa and helped popularise microscopy as a technique. On 9 October 1676, Leeuwenhoek reported the discovery of micro-organisms.

In 1893 August Köhler development a key technique for sample illumination, Köhler illumination, which is central to modern light microscopy. This method of sample illumination gives rise to extremely even lighting and overcomes many limitations of older techniques of sample illumination. Further developments in sample illumination came from Fritz Zernike in 1953 and George Nomarski 1955 for their development of phase contrast and differential interference contrast illumination which allow imaging of transparent samples.

Electron microscopy



A Flea as imaged using an electron microscope

In the early 1900s a significant alternative to light microscopy was developed, using electrons rather than light to generate the image. Ernst Ruska started development of the first electron microscope in 1931 which was the transmission electron microscope (TEM). The transmission electron microscope works on the same principle as an optical microscope but uses electrons in the place of light and electromagnets in the place of glass lenses. Use of electrons instead of light allows a much higher resolution.

Development of the transmission electron microscope was quickly followed in 1935 by the development of the scanning electron microscope by Max Knoll.

Electron microscopes quickly became popular, following the Second World War Ernst Ruska, working at Siemens developed the first commercial transmission electron microscope and major scientific conferences on electron microscopy started being held in the 1950s. In 1965 the first commercial scanning electron microscope was developed by Professor Sir Charles Oatley and his postgraduate student Gary Stewart and marketed by the Cambridge Instrument Company as the "Stereoscan".

Scanning probe microscopy

The 1980s saw the development of the first scanning probe microscopes. The first was the scanning tunneling microscope in 1981, developed by Gerd Binnig and Heinrich Rohrer. Whis was closely followed in 1986 with Gerd Binnig, Quate, and Gerber's invention of the atomic force microscope.

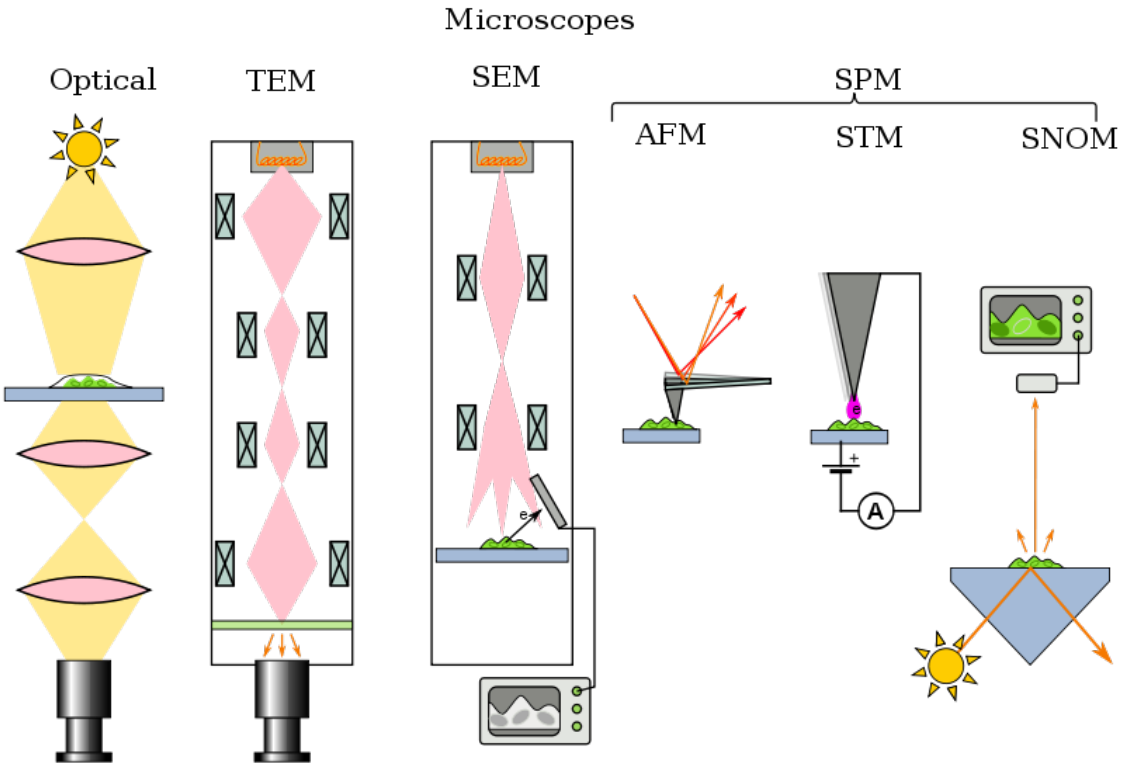
Fluorescence and light microscopy

The most recent developments in light microscope largely centre on the rise of fluorescence microscopy in biology. During the last decades of the 20th century, particularly in the post-genomic era, many techniques for fluorescent labeling of cellular structures were developed. The main groups of techniques are small chemical staining of cellular structures, for example DAPI to label DNA, use of antibodies conjugated to fluorescent reporters, see immunofluorescence, and fluorescent proteins, such as green fluorescent protein. These techniques use these different fluorophores for analysis of cell structure at a molecular level in both live and fixed samples.

The rise of fluorescence microscopy drove the development of a major modern microscope design, the confocal microscope. The principle was patented in 1957 by Marvin Minsky, although laser technology limited practical application of the technique. It was not until 1978 when Thomas and Christoph Cremer developed the first practical confocal laser scanning microscope and the technique rapidly gained popularity through the 1980s.

Much current research (in the early 21st century) on optical microscope techniques is focused on development of superresolution analysis of fluorescently labeled samples. Structured illumination can improve resolution by around two to four times and techniques like stimulated Emission Depletion microscopy are approaching the resolution of electron microscopes.

Types



Types of microscopes

Microscopes can be separated into several different classes. One grouping is based on what interacts with the sample to generate the image, i.e., light (optical microscopes), electrons (electron microscopes) or a probe (scanning probe microscopes). Alternatively microscopes can be classed on whether they analyse the sample via a scanning point (confocal optical microscopes, scanning electron microscopes and scanning probe microscopes) or analyse the sample all at once (wide field optical microscope and transmission electron microscopes).

The wide field optical microscope and transmission electron microscope use the theory of lenses (optics for light microscopes and electromagnet lenses for electron microscopes) in order to magnify the image generated by the passage of a wave through the sample, or reflected by the sample. The waves used are electromagnetic (in optical microscopes) or electron beams (in electron microscopes). Resolution in these microscopes is limited by the wavelength of the radiation used to image the sample, shorter wavelengths allow a higher resolution.

Scanning optical and electron microscopes, like the confocal microscope and scanning electron microscope, use lenses to focus a spot of light/electrons onto the sample then analyse the reflected and/or transmitted waves. The point is then scanned over the sample

to analyse a rectangular region. Magnification of the image is achieved by displaying the data from scanning a small sample area on a large screen. These microscopes have the same resolution limit as wide field optical and electron microscopes.

Scanning probe microscopes also analyse a single point in the sample and then scan the probe over a rectangular sample region to build up an image. As these microscopes do not use electromagnetic or electron radiation for imaging they are not subject to the same resolution limit as the optical and electron microscopes described above.

Optical

The most common type of microscope—and the first invented—is the optical microscope. This is an optical instrument containing one or more lenses producing an enlarged image of an sample placed in the focal plane. Optical microscopes have refractive glass and occasionally of plastic or quartz, to focus light into the eye or another light detector. Mirror-based optical microscopes operate in the same manner. Typical magnification of a light microscope, assuming visible range light, is up to 1500x with a theoretical resolution limit of around 0.2 micrometres or 200 nanometers. Specialized techniques (e.g., scanning confocal microscopy, Vertico SMI) may exceed this magnification but the resolution is diffraction limited. The use of shorter wavelengths of light, such as the ultraviolet, is one way to improve the spatial resolution of the optical microscope, as are devices such as the near-field scanning optical microscope.
Sarfus, a recent optical technique increases the sensitivity of standard optical microscope to a point it becomes possible to directly visualize nanometric films (down to 0.3 nanometer) and isolated nano-objects (down to 2 nm-diameter). The technique is based on the use of non-reflecting substrates for cross-polarized reflected light microscopy.

CBP Office of Field Operations agent checking the authenticity of a travel document at an international airport using a stereo microscope

Ultraviolet light enables the resolution of microscopic features, as well as to image samples that are transparent to the eye. Near infrared light images circuitry embedded in bonded silicon devices, as silicon is transparent in this region. Many wavelengths of light, ranging from the ultraviolet to the visible are used to excite fluorescence emission from objects for viewing by eye or with sensitive cameras.

Phase contrast microscopy is an optical microscopy illumination technique in which small phase shifts in the light passing through a transparent specimen are converted into amplitude or contrast changes in the image. A phase contrast microscope does not require staining to view the slide. This microscope made it possible to study the cell cycle.

The traditional optical microscope has recently been modified into a digital microscope, where, instead of directly viewing the object, a charge-coupled device (CCD) is used to record the image, which is then displayed on a computer monitor.

Electron

Three major variants of electron microscopes exist:

- Scanning electron microscope (SEM): looks at the surface of bulk objects by scanning the surface with a fine electron beam and measuring reflection. May also be used for spectroscopy.
- Transmission electron microscope (TEM): passes electrons completely through the sample, analogous to basic optical microscopy. This requires careful sample preparation, since electrons are scattered so strongly by most materials.This is a scientific device that allows people to see objects that could normally not be seen by the naked or unaided eye.
- Scanning Tunneling Microscope (STM): is a powerful technique for viewing surfaces at the atomic level.

The SEM and STM can also be considered examples of scanning probe microscopy.

Scanning probe

- AFM, atomic force microscopy
- BEEM, ballistic electron emission microscopy
- EFM, electrostatic force microscope
- ESTM electrochemical scanning tunneling microscope
- FMM, force modulation microscopy
- KPFM, kelvin probe force microscopy
- MFM, magnetic force microscopy
- MRFM, magnetic resonance force microscopy
- NSOM, near-field scanning optical microscopy (or SNOM, scanning near-field optical microscopy)
- PFM, Piezo Force Microscopy
- PSTM, photon scanning tunneling microscopy
- PTMS, photothermal microspectroscopy/microscopy
- SAP, scanning atom probe
- SCM, scanning capacitance microscopy
- SECM, scanning electrochemical microscopy
- SEM, scanning electron microscopy
- SGM, scanning gate microscopy
- SICM, scanning ion-conductance microscopy
- SPSM spin polarized scanning tunneling microscopy
- SThM, scanning thermal microscopy/annurev.matsci.29.1.505]
- STM, scanning tunneling microscopy
- SVM, scanning voltage microscopy
- SHPM, scanning Hall probe microscopy
- SSM, Scanning SQUID microscope

Of these techniques AFM and STM are the most commonly used followed by MFM and SNOM/NSOM.

Other

Replica of microscope by Van Leeuwenhoek

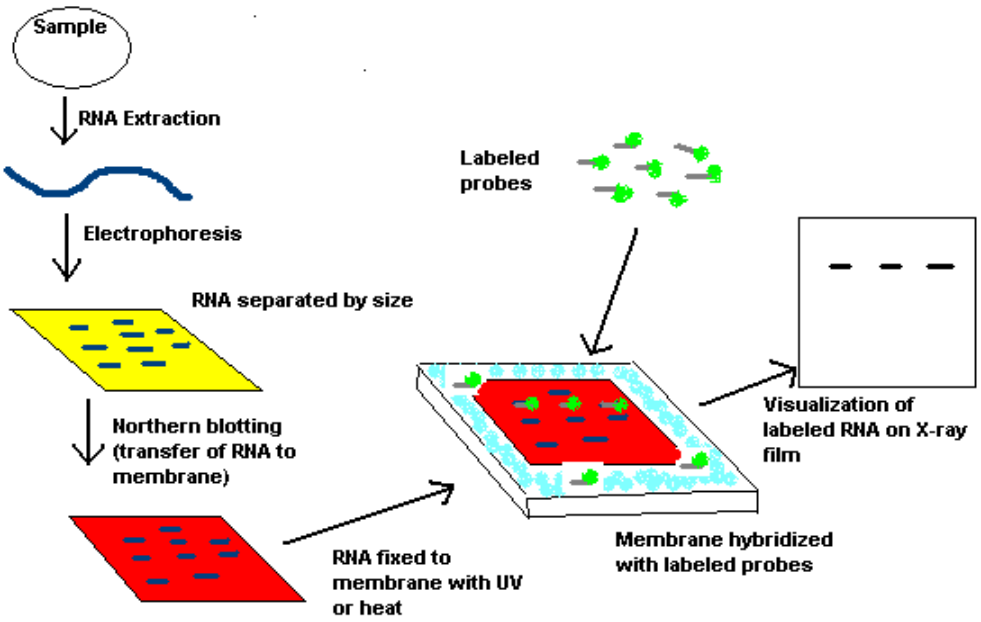
Different microscopes

Scanning acoustic microscopes use sound waves to measure variations in acoustic impedance. Similar to Sonar in principle, they are used for such jobs as detecting defects in the subsurfaces of materials including those found in integrated circuits.

Chapter- 5

Northern Blot

The **northern blot** is a technique used in molecular biology research to study gene expression by detection of RNA (or isolated mRNA) in a sample.



Flow diagram outlining the general procedure for RNA detection by northern blotting.

With northern blotting it is possible to observe cellular control over structure and function by determining the particular gene expression levels during differentiation, morphogenesis, as well as abnormal or diseased conditions. Northern blotting involves the use of electrophoresis to separate RNA samples by size, and detection with a

hybridization probe complementary to part of or the entire target sequence. The term 'northern blot' actually refers specifically to the capillary transfer of RNA from the electrophoresis gel to the blotting membrane, however the entire process is commonly referred to as northern blotting. The northern blot technique was developed in 1977 by James Alwine, David Kemp, and George Stark at Stanford University. Northern blotting takes its name from its similarity to the first blotting technique, the Southern blot, named for biologist Edwin Southern. The major difference is that RNA, rather than DNA, is analyzed in the Northern blot.

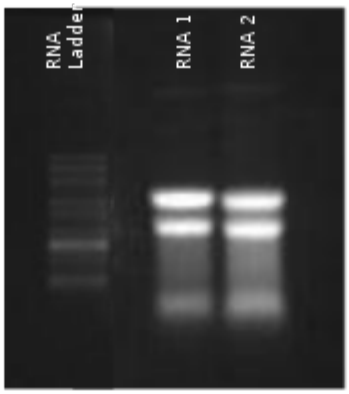
Procedure

A general blotting procedure starts with extraction of total RNA from a homogenized tissue sample. The mRNA can then be isolated through the use of oligo (dT) cellulose chromatography to maintain only those RNAs with a poly(A) tail. RNA samples are then separated by gel electrophoresis. Since the gels are fragile and the probes are unable to enter the matrix, the RNA samples, now separated by size, are transferred to a nylon membrane through a capillary or vacuum blotting system.

Capillary blotting system setup for the transfer of RNA from an electrophoresis gel to a blotting membrane.

A nylon membrane with a positive charge is the most effective for use in northern blotting since the negatively charged nucleic acids have a high affinity for them. The transfer buffer used for the blotting usually contains formamide because it lowers the annealing temperature of the probe-RNA interaction preventing RNA degradation by high temperatures. Once the RNA has been transferred to the membrane it is immobilized through covalent linkage to the membrane by UV light or heat. After a probe has been labeled, it is hybridized to the RNA on the membrane. Experimental conditions that can affect the efficiency and specificity of hybridization include ionic strength, viscosity, duplex length, mismatched base pairs, and base composition. The membrane is washed to ensure that the probe has bound specifically and to avoid background signals from arising. The hybrid signals are then detected by X-ray film and can be quantified by densitometry. To create controls for comparison in a northern blot, samples not displaying the gene product of interest can be used after determination by microarrays or RT-PCR.

Gels



Formaldehyde gel (1%) with RNA samples run at 100V for 1 hour in 1× MOPS buffe

RNA run on a formaldehyde agarose gel to highlight the 28S (top band) and 18S (lower band) ribosomal subunits.

The RNA samples are most commonly separated on agarose gels containing formaldehyde as a denaturing agent for the RNA to limit secondary structure. The gels can be stained with ethidium bromide (EtBr) and viewed under UV light to observe the quality and quantity of RNA before blotting. Polyacrylamide gel electrophoeresis with urea can also be used in RNA separation but it is most commonly used for fragmented RNA or microRNAs. An RNA ladder is often run alongside the samples on an electrophoresis gel to observe the size of fragments obtained but in total RNA samples the ribosomal subunits can act as size markers. Since the large ribosomal subunit is 28S (approximately 5kb) and the small ribosomal subunit is 18S (approximately 2kb) two

prominent bands will appear on the gel, the larger at close to twice the intensity of the smaller.

Probes

Probes for northern blotting are composed of nucleic acids with a complementary sequence to all or part of the RNA of interest, they can be DNA, RNA, or oligonucleotides with a minimum of 25 complementary bases to the target sequence. RNA probes (riboprobes) that are transcribed in vitro are able to withstand more rigorous washing steps preventing some of the background noise. Commonly cDNA is created with labelled primers for the RNA sequence of interest to act as the probe in the northern blot. The probes need to be labelled either with radioactive isotopes (^{32}P) or with chemiluminescence in which alkaline phosphatase or horseradish peroxidase breakdown chemiluminescent substrates producing a detectable emission of light. The chemiluminescent labelling can occur in two ways: either the probe is attached to the enzyme, or the probe is labelled with a ligand (e.g. biotin) for which the antibody (e.g. avidin or streptavidin) is attached to the enzyme. X-ray film can detect both the radioactive and chemiluminescent signals and many researchers prefer the chemiluminescent signals because they are faster, more sensitive, and reduce the health hazards that go along with radioactive labels. The same membrane can be probed up to five times without a significant loss of the target RNA.

Applications

Northern blotting allows one to observe a particular gene's expression pattern between tissues, organs, developmental stages, environmental stress levels, pathogen infection, and over the course of treatment. The technique has been used to show overexpression of oncogenes and downregulation of tumor-suppressor genes in cancerous cells when compared to 'normal' tissue, as well as the gene expression in the rejection of transplanted organs. If an upregulated gene is observed by an abundance of mRNA on the northern blot the sample can then be sequenced to determine if the gene is known to researchers or if it is a novel finding. The expression patterns obtained under given conditions can provide insight into the function of that gene. Since the RNA is first separated by size, if only one probe type is used variance in the level of each band on the membrane can provide insight into the size of the product, suggesting alternative splice products of the same gene or repetitive sequence motifs. The variance in size of a gene product can also indicate deletions or errors in transcript processing, by altering the probe target used along the known sequence it is possible to determine which region of the RNA is missing.

BlotBase is an online database publishing northern blots. BlotBase has over 700 published northern blots of human and mouse samples, in over 650 genes across more than 25 different tissue types. Northern blots can be searched by a blot ID, paper reference, gene identifier, or by tissue. The results of a search provide the blot ID, species, tissue, gene, expression level, blot image (if available), and links to the publication that the work originated from. This new database provides sharing of

information between members of the science community that was not previously seen in northern blotting as it was in sequence analysis, genome determination, protein structure, etc.

Disadvantages and Advantages

Analysis of gene expression can be done by several different methods including RT-PCR, RNase protection assays, microarrays, serial analysis of gene expression (SAGE), as well as northern blotting. Microarrays are quite commonly used and are usually consistent with data obtained from northern blots; however, at times northern blotting is able to detect small changes in gene expression that microarrays cannot. The advantage that microarrays have over northern blots is that thousands of genes can be visualized at a time, while northern blotting is usually looking at one or a small number of genes.

A problem in northern blotting is often sample degradation by RNases (both endogenous to the sample and through environmental contamination), which can be avoided by proper sterilization of glassware and the use of RNase inhibitors such as DEPC (diethylpyrocarbonate). The chemicals used in most northern blots can be a risk to the researcher, since formaldehyde, radioactive material, ethidium bromide, DEPC, and UV light are all harmful under certain exposures. Compared to RT-PCR, northern blotting has a low sensitivity, but it also has a high specificity which is important to reduce false positive results.

The advantages of using northern blotting include the detection of RNA size, the observation of alternate splice products, the use of probes with partial homology, the quality and quantity of RNA can be measured on the gel prior to blotting, and the membranes can be stored and reprobed for years after blotting.

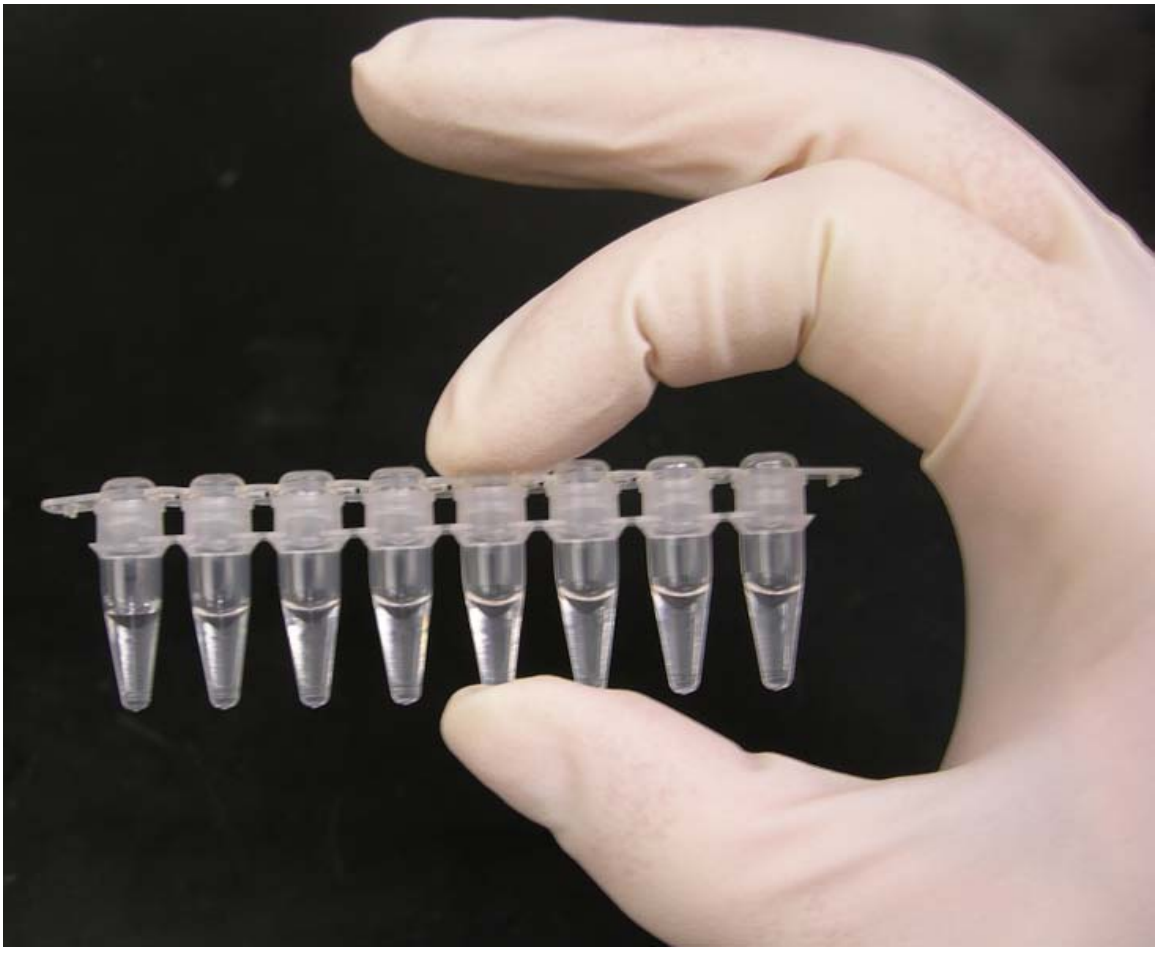
Reverse northern blot

A variant of the procedure known as the reverse northern blot is occasionally used. In this procedure, the substrate nucleic acid (that is affixed to the membrane) is a collection of isolated DNA fragments, and the probe is RNA extracted from a tissue and radioactively labelled.

The use of DNA microarrays that have come into widespread use in the late 1990s and early 2000s is more akin to the reverse procedure, in that they involve the use of isolated DNA fragments affixed to a substrate, and hybridization with a probe made from cellular RNA. Thus the reverse procedure, though originally uncommon, enabled northern analysis to evolve into gene expression profiling, in which many (possibly all) of the genes in an organism may have their expression monitored.

Chapter- 6

Polymerase Chain Reaction



A strip of eight PCR tubes, each containing a 100 µl reaction mixture

The **polymerase chain reaction** (**PCR**) is a scientific technique in molecular biology to amplify a single or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.

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 forensic sciences and paternity testing); and the detection and diagnosis of infectious diseases. In 1993, Mullis was awarded the Nobel Prize in Chemistry for his work on PCR.

The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers (short DNA fragments) containing sequences complementary to the target region along with a DNA polymerase (after which the method is named) are key components to enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. PCR 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, an enzyme originally isolated from the bacterium *Thermus aquaticus*. This DNA polymerase enzymatically assembles a new DNA strand from DNA building blocks, the nucleotides, by using single-stranded DNA as a template and DNA oligonucleotides (also called DNA primers), which are 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 thermal cycling steps are necessary first to physically separate the two strands in a DNA double helix at a high temperature in a process called DNA melting. At a lower temperature, each strand is then used as the template in DNA synthesis by the DNA polymerase to selectively amplify the target DNA. The selectivity of PCR results from the use of primers that are complementary to the DNA region targeted for amplification under specific thermal cycling conditions.

PCR principles and procedure

Figure 1a: A thermal cycler for PCR

Figure 1b: An older model three-temperature thermal cycler for PCR

PCR is used to amplify a specific region of a DNA strand (the DNA target). 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.

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

- *DNA template* that contains the DNA region (target) to be amplified.
- Two *primers* that are complementary to the 3' (three prime) ends of each of the sense and anti-sense strand of the DNA target.

- *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
- *Monovalent cation* potassium ions.

The PCR is commonly carried out in a reaction volume of 10–200 µl in small reaction tubes (0.2–0.5 ml volumes) in a thermal cycler. The thermal cycler heats and cools the reaction tubes to achieve the temperatures required at each step of the reaction. Many modern thermal cyclers make use of the Peltier effect which permits both heating and cooling of the block holding the PCR tubes simply by reversing the electric current. 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

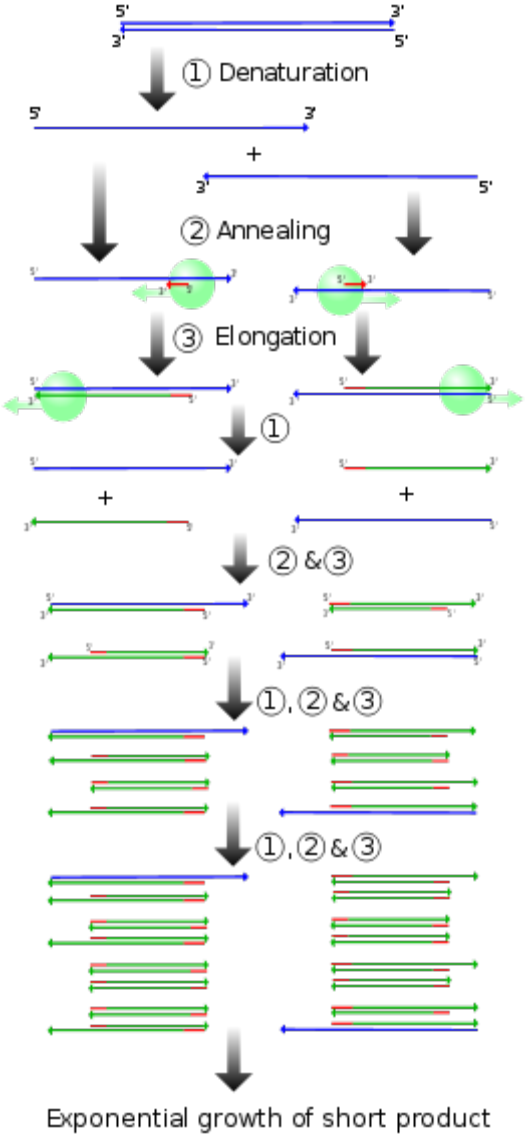


Figure 2: Schematic drawing of the PCR cycle. **(1) Denaturing at 94–96 °C. (2) Annealing at ~65 °C (3) Elongation at 72 °C**. Four cycles are shown here. The blue lines represent the DNA template to which primers (red arrows) anneal that are extended by the DNA polymerase (light green circles), to give shorter DNA products (green lines), which themselves are used as templates as PCR progresses.

Typically, PCR consists of a series of 20-40 repeated temperature changes, called cycles, with each cycle commonly consisting of 2-3 discrete temperature steps, usually three (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.

- *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.

- *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 DNA melting of the DNA template by disrupting the hydrogen bonds between complementary bases, yielding single-stranded DNA molecules.

- *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, 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 dNTPs 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 per minute. Under optimum conditions, i.e., if there are no limitations due to limiting substrates or reagents, at each extension step, the amount of DNA target is doubled, leading to exponential (geometric) amplification of the specific DNA fragment.

- *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.

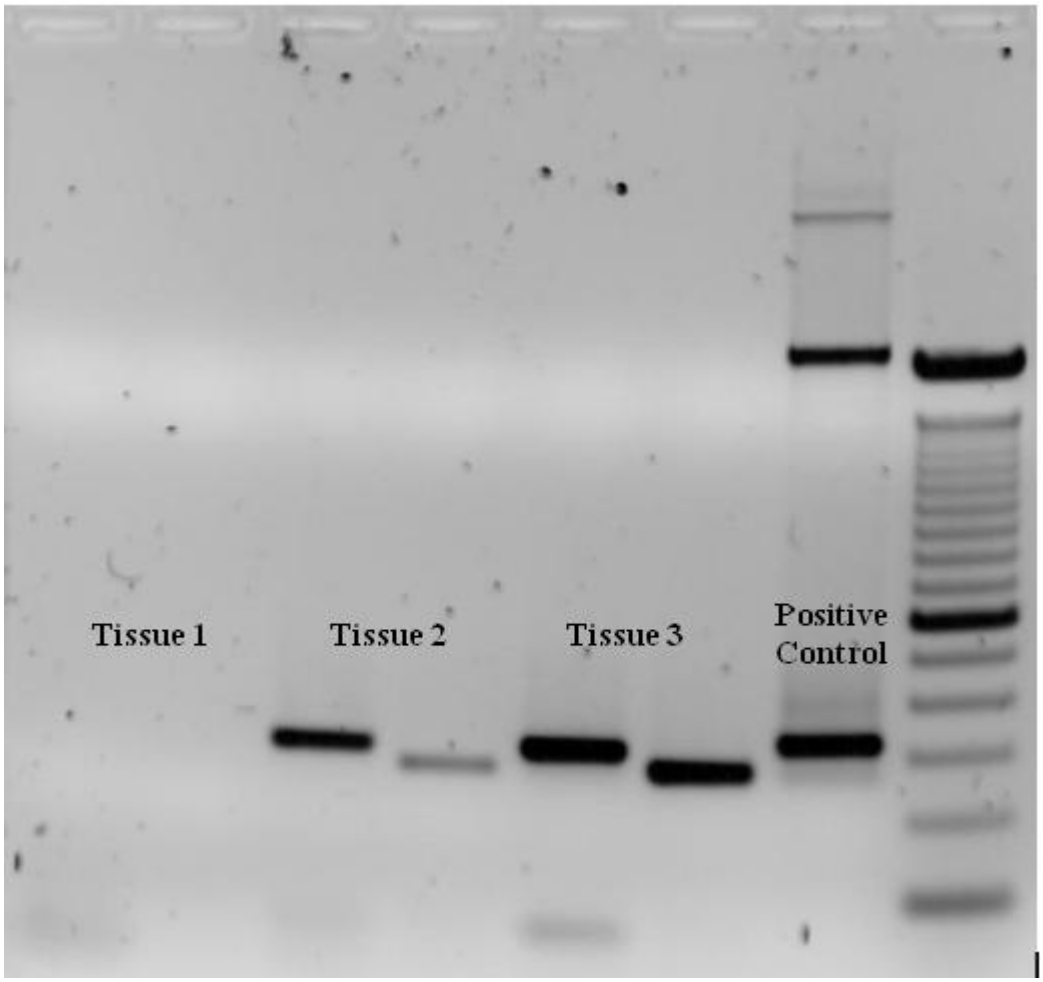


Figure 3: Ethidium bromide-stained PCR products after gel electrophoresis. Two sets of primers were used to amplify a target sequence from three different tissue samples. No amplification is present in sample #1; DNA bands in sample #2 and #3 indicate successful amplification of the target sequence. 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 the amplimer or amplicon), 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 (a molecular weight marker), which contains DNA fragments of known size, run on the gel alongside the PCR products.

PCR stages

The PCR process can be divided into three stages:

Exponential amplification: At every cycle, the amount of product is doubled (assuming 100% reaction efficiency). The reaction is very sensitive: only minute quantities of DNA need to be present.

Levelling off stage: The reaction slows as the DNA polymerase loses activity and as consumption of reagents such as dNTPs and primers causes them to become limiting.

Plateau: No more product accumulates due to exhaustion of reagents and enzyme.

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 mixtures from potential DNA contaminants. This usually involves spatial separation of PCR-setup areas from areas for analysis or purification of PCR products, use of disposable plasticware, 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. Addition of reagents, such as formamide, in buffer systems may increase the specificity and yield of PCR.

Application of PCR

Selective DNA isolation

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 generating hybridization probes for Southern or northern hybridization and DNA cloning, which require larger 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. Bacterial colonies (E. coli) can be rapidly screened by PCR for correct DNA vector constructs. PCR may also be used for genetic fingerprinting; a forensic technique used to identify a person or organism by comparing experimental DNAs through different PCR-based methods.

Some PCR 'fingerprints' methods have high discriminative power and can be used to identify genetic relationships between individuals, such as parent-child or between

siblings, and are used in paternity testing (Fig. 4). This technique may also be used to determine evolutionary relationships among organisms.

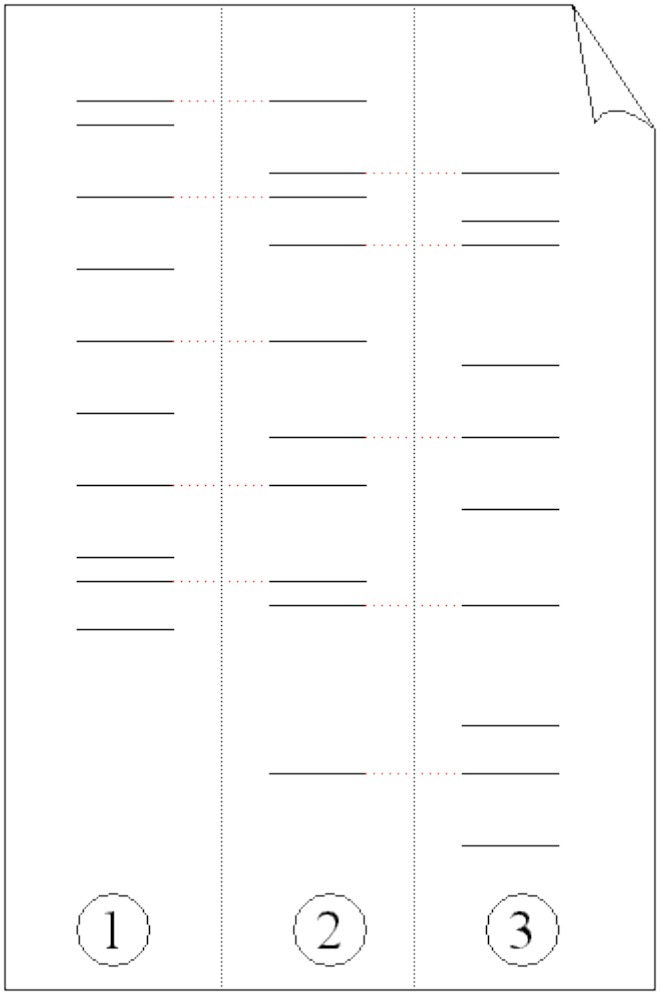


Figure 4: Electrophoresis of PCR-amplified DNA fragments. (1) Father. (2) Child. (3) Mother. The child has inherited some, but not all of the fingerprint of each of its parents, giving it a new, unique fingerprint.

Amplification and quantification 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 tens of 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.

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.

PCR in diagnosis of diseases

PCR permits early diagnosis of malignant diseases such as leukemia and lymphomas, which is currently the highest developed in cancer research and is already being used routinely. PCR assays can be performed directly on genomic DNA samples to detect translocation-specific malignant cells at a sensitivity which is at least 10,000 fold higher than other methods.

PCR also permits identification of non-cultivatable or slow-growing microorganisms such as mycobacteria, anaerobic bacteria, or viruses from tissue culture assays and animal models. The basis for PCR diagnostic applications in microbiology is the detection of infectious agents and the discrimination of non-pathogenic from pathogenic strains by virtue of specific genes.

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.

Variations on the basic PCR technique

- *Allele-specific PCR*: a diagnostic or cloning technique which is based on 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 an SNP-specific primer signals presence of the specific SNP in a sequence.

- *Assembly PCR* or *Polymerase Cycling Assembly (PCA)*: 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.

- *Asymmetric PCR*: preferentially amplifies one DNA strand in a double-stranded DNA template. It is used in sequencing and hybridization probing where amplification of only one of the two complementary strands is required. PCR is carried out as usual, but with a great excess of the primer for the strand targeted

for amplification. Because of the slow (arithmetic) amplification later in the reaction after the limiting primer has been used up, extra cycles of PCR are required. 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.

- *Helicase-dependent amplification*: 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.

- *Hot-start PCR*: a technique that reduces non-specific amplification during the initial set up stages of the PCR. It may be performed manually by heating the reaction components to the melting temperature (e.g., 95°C) before adding the polymerase. Specialized enzyme systems have been developed that inhibit the polymerase's activity at ambient temperature, either by the binding of an antibody 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 PCR* (ISSR): a PCR method for DNA fingerprinting that amplifies regions between simple sequence repeats to produce a unique fingerprint of amplified fragment lengths.

- *Inverse PCR*: is commonly used to identify the flanking sequences around genomic inserts. It involves a series of DNA digestions and self ligation, resulting in known sequences at either end of the unknown sequence.

- *Ligation-mediated PCR*: 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.

- *Methylation-specific PCR* (MSP): developed by Stephen Baylin and Jim Herman at the Johns Hopkins School of Medicine, 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 PCRs 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.

- *Miniprimer PCR*: uses a thermostable polymerase (S-Tbr) that can extend from short primers ("smalligos") as short as 9 or 10 nucleotides. This method permits PCR targeting to smaller primer binding regions, and is used to amplify conserved DNA sequences, such as the 16S (or eukaryotic 18S) rRNA gene.

- *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.

- *Multiplex-PCR*: consists of multiple primer sets within a single PCR mixture to produce amplicons of varying sizes that are 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 used in two successive PCRs. 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 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.

- *Overlap-extension PCR*: a genetic engineering technique allowing the construction of a DNA sequence with an alteration inserted beyond the limit of the longest practical primer length.

- *Quantitative PCR (Q-PCR)*: used to measure the quantity of a PCR product (commonly in real-time). It quantitatively measures 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. *Quantitative real-time PCR* has a very high degree of precision. QRT-PCR methods use fluorescent dyes, such as Sybr Green, EvaGreen or fluorophore-containing DNA probes, such as TaqMan, to measure the amount of amplified product in real time. It is also sometimes abbreviated to RT-PCR (*R*eal *T*ime PCR) or RQ-PCR. QRT-PCR or RTQ-PCR are more appropriate contractions, since RT-PCR commonly refers to reverse transcription PCR, often used in conjunction with Q-PCR.

- *R*everse *T*ranscription PCR (RT-PCR): for amplifying DNA from RNA. Reverse transcriptase reverse transcribes RNA into cDNA, which is then amplified by

PCR. 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. If the genomic DNA sequence of a gene is known, RT-PCR can be used 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 RACE-PCR (*Rapid Amplification of cDNA Ends*).

- *Solid Phase PCR*: encompasses multiple meanings, including Polony Amplification (where PCR colonies are derived in a gel matrix, for example), Bridge PCR (primers are covalently linked to a solid-support surface), conventional Solid Phase PCR (where Asymmetric PCR is applied in the presence of solid support bearing primer with sequence matching one of the aqueous primers) and Enhanced Solid Phase PCR (where conventional Solid Phase PCR can be improved by employing high Tm and nested solid support primer with optional application of a thermal 'step' to favour solid support priming).

- *Thermal asymmetric interlaced PCR (TAIL-PCR)*: for isolation of an 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.

- *Touchdown PCR* (*Step-down 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.

- *PAN-AC*: uses isothermal conditions for amplification, and may be used in living cells.

- *Universal Fast Walking*: for genome walking and genetic fingerprinting using a more specific 'two-sided' PCR than conventional 'one-sided' approaches (using only one gene-specific primer and one general primer - which can lead to artefactual 'noise') by virtue of a mechanism involving lariat structure formation. Streamlined derivatives of UFW are LaNe RAGE (lariat-dependent nested PCR for rapid amplification of genomic DNA ends), 5'RACE LaNe and 3'RACE LaNe.

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*. 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.

At the core of the PCR method is the use of a suitable DNA polymerase able to withstand the high temperatures of >90 °C (194 °F) required for separation of the two DNA strands in the DNA double helix after each replication cycle. The DNA polymerases initially employed for in vitro experiments presaging PCR were unable to withstand these high temperatures. So the early procedures for DNA replication were very inefficient, time consuming, and required large amounts of DNA polymerase and continual handling throughout the process.

The discovery in 1976 of Taq polymerase — a DNA polymerase purified from the thermophilic bacterium, *Thermus aquaticus*, which naturally lives in hot (50 to 80 °C (122 to 176 °F)) environments such as hot springs — paved the way for dramatic improvements of the PCR method. The DNA polymerase isolated from *T. aquaticus* is stable at high temperatures remaining active even after DNA denaturation, thus obviating the need to add new DNA polymerase after each cycle. This allowed an automated thermocycler-based process for DNA amplification.

When Mullis developed the PCR in 1983, he 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. 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 DNA polymerase. In *Scientific American*, Mullis summarized the procedure: "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." He was awarded the Nobel Prize in Chemistry in 1993 for his invention, 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.

Patent wars

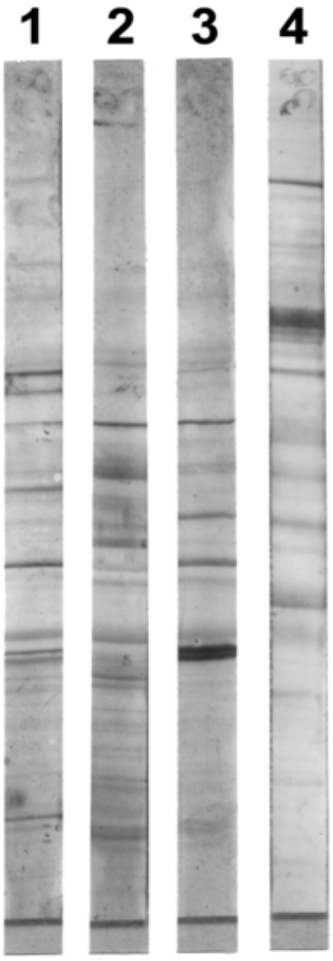
The PCR technique was patented by Kary Mullis and assigned to Cetus Corporation, where Mullis worked when he invented the technique in 1983. The *Taq* polymerase enzyme was also covered by patents. There have been several high-profile lawsuits related to the technique, including an unsuccessful lawsuit 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. The legal arguments have extended beyond the lives of the original PCR and Taq polymerase patents, which expired on March 28, 2005.

Chapter- 7

Western Blot and Southern Blot

Western blot



Western blot analysis of proteins separated by SDS-PAGE.

The **Western blot** (alternatively, **protein immunoblot**) is an extremely useful analytical technique used to detect specific proteins in the given sample of tissue homogenate or extract. It uses gel electrophoresis to separate native or denatured proteins by the length of the polypeptide (denaturing conditions) or by the 3-D structure of the protein (native/ non-denaturing conditions). The proteins are then transferred to a membrane (typically nitrocellulose or PVDF), where they are probed (detected) using antibodies specific to the target protein.

There are now many reagent companies that specialize in providing antibodies (both monoclonal and polyclonal antibodies) against tens of thousands of different proteins. Commercial antibodies can be expensive, although the unbound antibody can be reused between experiments. This method is used in the fields of molecular biology, biochemistry, immunogenetics and other molecular biology disciplines.

Other related techniques include using antibodies to detect proteins in tissues and cells by immunostaining and enzyme-linked immunosorbent assay (ELISA).

The method originated from the laboratory of George Stark at Stanford. The name *Western blot* was given to the technique by W. Neal Burnette and is a play on the name Southern blot, a technique for DNA detection developed earlier by Edwin Southern. Detection of RNA is termed northern blotting and the detection of post-translational modification of protein is termed eastern blotting.

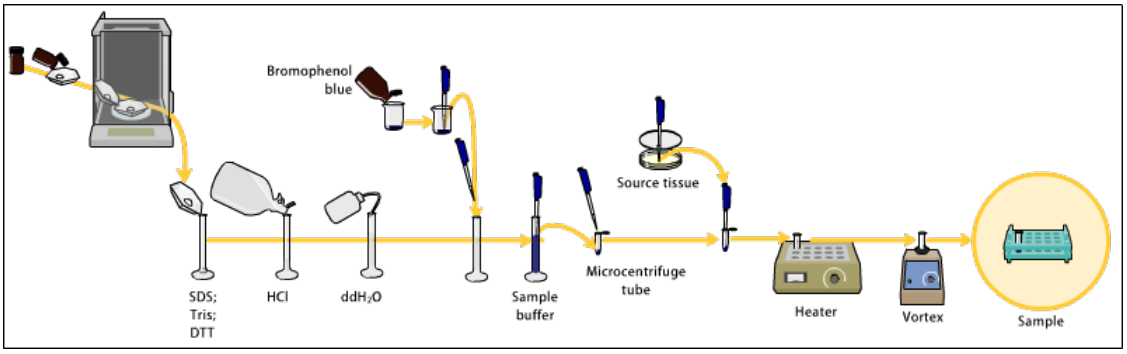
Steps in a Western blot

Tissue preparation

Samples may be taken from whole tissue or from cell culture. In most cases, solid tissues are first broken down mechanically using a blender (for larger sample volumes), using a homogenizer (smaller volumes), or by sonication. Cells may also be broken open by one of the above mechanical methods. However, bacteria, virus or environmental samples can be the source of protein and thus Western blotting is not restricted to cellular studies only.

Assorted detergents, salts, and buffers may be employed to encourage lysis of cells and to solubilize proteins. Protease and phosphatase inhibitors are often added to prevent the digestion of the sample by its own enzymes. Tissue preparation is often done at cold temperatures to avoid protein denaturing and degradation.

A combination of biochemical and mechanical techniques – including various types of filtration and centrifugation – can be used to separate different cell compartments and organelles.

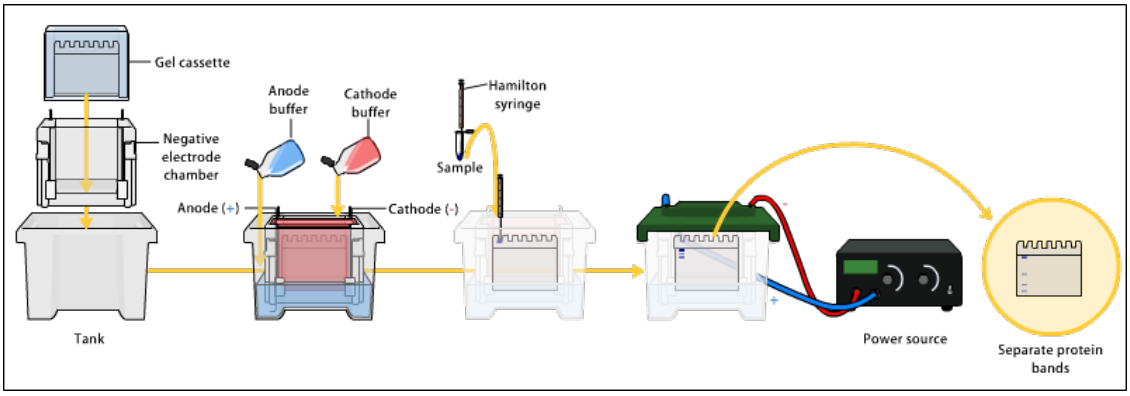


Gel electrophoresis

The proteins of the sample are separated using gel electrophoresis. Separation of proteins may be by isoelectric point (pI), molecular weight, electric charge, or a combination of these factors. The nature of the separation depends on the treatment of the sample and the nature of the gel. This is a very useful way to determine a protein.

By far the most common type of gel electrophoresis employs polyacrylamide gels and buffers loaded with sodium dodecyl sulfate (SDS). SDS-PAGE (SDS polyacrylamide gel electrophoresis) maintains polypeptides in a denatured state once they have been treated with strong reducing agents to remove secondary and tertiary structure (e.g. disulfide bonds [S-S] to sulfhydryl groups [SH and SH]) and thus allows separation of proteins by their molecular weight. Sampled proteins become covered in the negatively charged SDS and move to the positively charged electrode through the acrylamide mesh of the gel. Smaller proteins migrate faster through this mesh and the proteins are thus separated according to size (usually measured in kilodaltons, kDa). The concentration of acrylamide determines the resolution of the gel - the greater the acrylamide concentration the better the resolution of lower molecular weight proteins. The lower the acrylamide concentration the better the resolution of higher molecular weight proteins. Proteins travel only in one dimension along the gel for most blots.

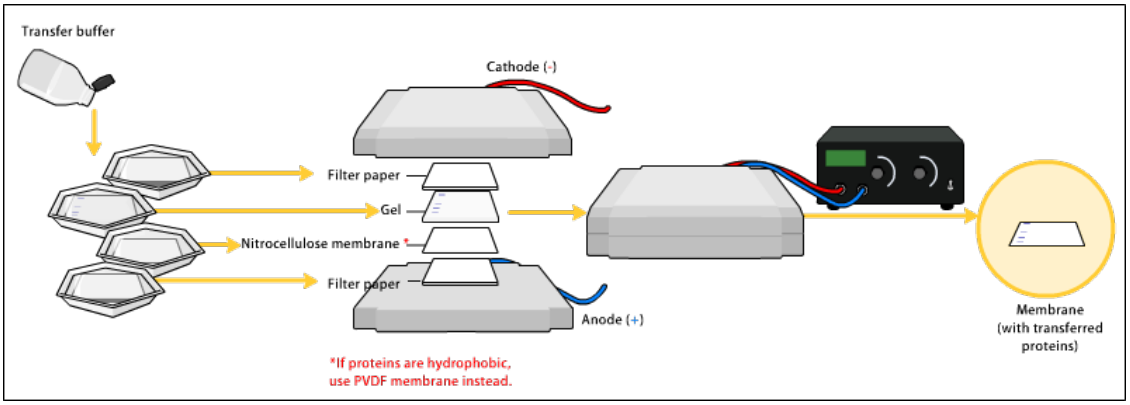
Samples are loaded into *wells* in the gel. One lane is usually reserved for a *marker* or *ladder*, a commercially available mixture of proteins having defined molecular weights, typically stained so as to form visible, coloured bands. When voltage is applied along the gel, proteins migrate into it at different speeds. These different rates of advancement (different *electrophoretic mobilities*) separate into *bands* within each *lane*.



It is also possible to use a two-dimensional (2-D) gel which spreads the proteins from a single sample out in two dimensions. Proteins are separated according to isoelectric point (pH at which they have neutral net charge) in the first dimension, and according to their molecular weight in the second dimension.

Transfer

In order to make the proteins accessible to antibody detection, they are moved from within the gel onto a membrane made of *nitrocellulose or polyvinylidene difluoride (PVDF*). The membrane is placed on top of the gel, and a stack of filter papers placed on top of that. The entire stack is placed in a buffer solution which moves up the paper by capillary action, bringing the proteins with it. Another method for transferring the proteins is called electroblotting and uses an electric current to pull proteins from the gel into the PVDF or nitrocellulose membrane. The protein move from within the gel onto the membrane while maintaining the organization they had within the gel. As a result of this "blotting" process, the proteins are exposed on a thin surface layer for detection. Both varieties of membrane are chosen for their non-specific protein binding properties (i.e. binds all proteins equally well). Protein binding is based upon hydrophobic interactions, as well as charged interactions between the membrane and protein. Nitrocellulose membranes are cheaper than PVDF, but are far more fragile and do not stand up well to repeated probings.



The uniformity and overall effectiveness of transfer of protein from the gel to the membrane can be checked by staining the membrane with Coomassie Brilliant Blue or Ponceau S dyes. Ponceau S is the more common of the two, due to Ponceau S's higher sensitivity and its water solubility makes it easier to subsequently destain and probe the membrane as described below.

Blocking

Since the membrane has been chosen for its ability to bind protein and as both antibodies and the target are proteins, steps must be taken to prevent interactions between the membrane and the antibody used for detection of the target protein. Blocking of non-specific binding is achieved by placing the membrane in a dilute solution of protein - typically 3-5% Bovine serum albumin (BSA) or non-fat dry milk (both are inexpensive) in Tris-Buffered Saline (TBS), with a minute percentage of detergent such as Tween 20 or Triton X-100. The protein in the dilute solution attaches to the membrane in all places where the target proteins have not attached. Thus, when the antibody is added, there is no room on the membrane for it to attach other than on the binding sites of the specific target protein. This reduces "noise" in the final product of the Western blot, leading to clearer results, and eliminates false positives.

Detection

During the detection process the membrane is "probed" for the protein of interest with a modified antibody which is linked to a reporter enzyme, which when exposed to an appropriate substrate drives a colourimetric reaction and produces a colour. For a variety of reasons, this traditionally takes place in a two-step process, although there are now one-step detection methods available for certain applications.

Two steps

- Primary antibody

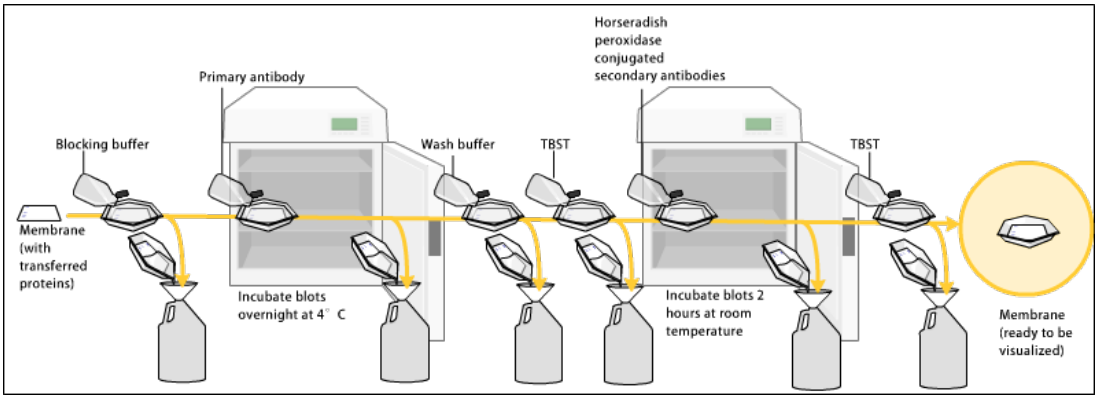
Antibodies are generated when a host species or immune cell culture is exposed to the protein of interest (or a part thereof). Normally, this is part of the immune response, whereas here they are harvested and used as sensitive and specific detection tools that bind the protein directly.

After blocking, a dilute solution of primary antibody (generally between 0.5 and 5 micrograms/mL) is incubated with the membrane under gentle agitation. Typically, the solution is composed of buffered saline solution with a small percentage of detergent, and sometimes with powdered milk or BSA. The antibody solution and the membrane can be sealed and incubated together for anywhere from 30 minutes to overnight. It can also be incubated at different temperatures, with warmer temperatures being associated with more binding, both specific (to the target protein, the "signal") and non-specific ("noise").

- Secondary antibody

After rinsing the membrane to remove unbound primary antibody, the membrane is exposed to another antibody, directed at a species-specific portion of the primary antibody. Antibodies come from animal sources (or animal sourced hybridoma cultures); an anti-mouse secondary will bind to almost any mouse-sourced primary antibody, which allows some cost savings by allowing an entire lab to share a single source of mass-produced antibody, and provides far more consistent results. This is known as a secondary antibody, and due to its targeting properties, tends to be referred to as "anti-mouse," "anti-goat," etc. The secondary antibody is usually linked to biotin or to a reporter enzyme such as alkaline phosphatase or horseradish peroxidase. This means that several secondary antibodies will bind to one primary antibody and enhance the signal.

Most commonly, a horseradish peroxidase-linked secondary is used to cleave a chemiluminescent agent, and the reaction product produces luminescence in proportion to the amount of protein. A sensitive sheet of photographic film is placed against the membrane, and exposure to the light from the reaction creates an image of the antibodies bound to the blot. A cheaper but less sensitive approach utilizes a 4-chloronaphthol stain with 1% hydrogen peroxide; reaction of peroxide radicals with 4-chloronaphthol produces a dark brown stain that can be photographed without using specialized photographic film.



As with the ELISPOT and ELISA procedures, the enzyme can be provided with a substrate molecule that will be converted by the enzyme to a colored reaction product that will be visible on the membrane.

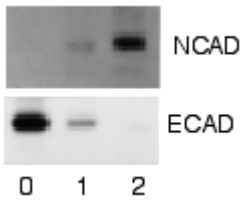
Another method of secondary antibody detection utilizes a near-infrared (NIR) fluorophore-linked antibody. Light produced from the excitation of a fluorescent dye is static, making fluorescent detection a more precise and accurate measure of the difference in signal produced by labeled antibodies bound to proteins on a Western blot. Proteins can be accurately quantified because the signal generated by the different amounts of proteins on the membranes is measured in a static state, as compared to chemiluminescence, in which light is measured in a dynamic state.

A third alternative is to use a radioactive label rather than an enzyme coupled to the secondary antibody, such as labeling an antibody-binding protein like *Staphylococcus* Protein A or Streptavidin with a radioactive isotope of iodine. Since other methods are

safer, quicker, and cheaper, this method is now rarely used; however, an advantage of this approach is the sensitivity of auto-radiography based imaging, which enables highly accurate protein quantification when combined with optical software (e.g. Optiquant).

One step

Historically, the probing process was performed in two steps because of the relative ease of producing primary and secondary antibodies in separate processes. This gives researchers and corporations huge advantages in terms of flexibility, and adds an amplification step to the detection process. Given the advent of high-throughput protein analysis and lower limits of detection, however, there has been interest in developing one-step probing systems that would allow the process to occur faster and with less consumables. This requires a probe antibody which both recognizes the protein of interest and contains a detectable label, probes which are often available for known protein tags. The primary probe is incubated with the membrane in a manner similar to that for the primary antibody in a two-step process, and then is ready for direct detection after a series of wash steps.



Western blot using radioactive detection system

Analysis

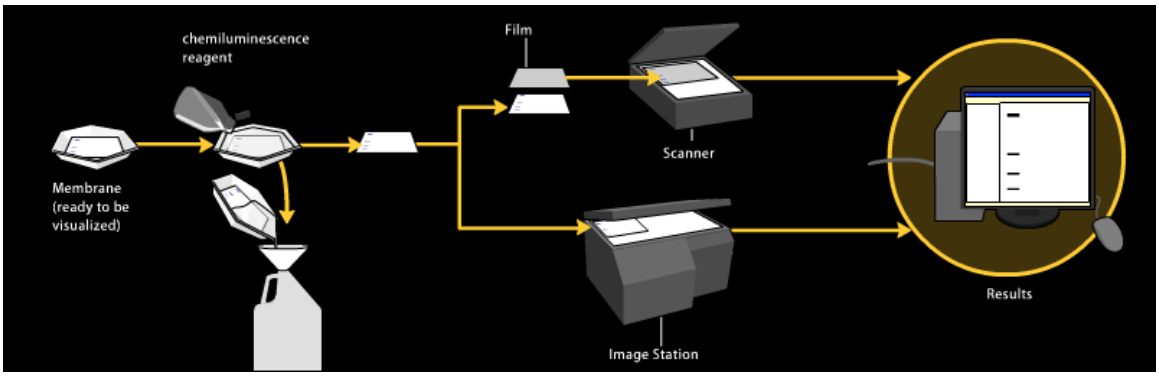
After the unbound probes are washed away, the Western blot is ready for detection of the probes that are labeled and bound to the protein of interest. In practical terms, not all Westerns reveal protein only at one band in a membrane. Size approximations are taken by comparing the stained bands to that of the marker or ladder loaded during electrophoresis. The process is repeated for a structural protein, such as actin or tubulin, that should not change between samples. The amount of target protein is indexed to the structural protein to control between groups. This practice ensures correction for the amount of total protein on the membrane in case of errors or incomplete transfers.

Colorimetric detection

The colorimetric detection method depends on incubation of the Western blot with a substrate that reacts with the reporter enzyme (such as peroxidase) that is bound to the secondary antibody. This converts the soluble dye into an insoluble form of a different color that precipitates next to the enzyme and thereby stains the membrane. Development of the blot is then stopped by washing away the soluble dye. Protein levels are evaluated through densitometry (how intense the stain is) or spectrophotometry.

Chemiluminescent detection

Chemiluminescent detection methods depend on incubation of the Western blot with a substrate that will luminesce when exposed to the reporter on the secondary antibody. The light is then detected by photographic film, and more recently by CCD cameras which capture a digital image of the Western blot. The image is analysed by densitometry, which evaluates the relative amount of protein staining and quantifies the results in terms of optical density. Newer software allows further data analysis such as molecular weight analysis if appropriate standards are used.



Radioactive detection

Radioactive labels do not require enzyme substrates, but rather allow the placement of medical X-ray film directly against the Western blot which develops as it is exposed to the label and creates dark regions which correspond to the protein bands of interest. The importance of radioactive detections methods is declining, because it is very expensive, health and safety risks are high, and ECL (enhanced chemiluminescence) provides a useful alternative.

Fluorescent detection

The fluorescently labeled probe is excited by light and the emission of the excitation is then detected by a photosensor such as CCD camera equipped with appropriate emission filters which captures a digital image of the Western blot and allows further data analysis such as molecular weight analysis and a quantitative Western blot analysis. Fluorescence is considered to be among the most sensitive detection methods for blotting analysis.

Secondary probing

One major difference between nitrocellulose and PVDF membranes relates to the ability of each to support "stripping" antibodies off and reusing the membrane for subsequent antibody probes. While there are well-established protocols available for stripping nitrocellulose membranes, the sturdier PVDF allows for easier stripping, and for more reuse before background noise limits experiments. Another difference is that, unlike

nitrocellulose, PVDF must be soaked in 95% ethanol, isopropanol or methanol before use. PVDF membranes also tend to be thicker and more resistant to damage during use.

2-D gel electrophoresis

2-dimensional SDS-PAGE uses the principles and techniques outlined above. 2-D SDS-PAGE, as the name suggests, involves the migration of polypeptides in 2 dimensions. For example, in the first dimension polypeptides are separated according to isoelectric point, while in the second dimension polypeptides are separated according to their molecular weight. The isoelectric point of a given protein is determined by the relative number of positively (e.g. lysine and arginine) and negatively (e.g. glutamate and aspartate) charged amino acids, with negatively charged amino acids contributing to a high isoelectric point and positively charged amino acids contributing to a low isoelectric point. Samples could also be separated first under nonreducing conditions using SDS-PAGE and under reducing conditions in the second dimension, which breaks apart disulfide bonds that hold subunits together. SDS-PAGE might also be coupled with urea-PAGE for a 2-dimensional gel.

In principle, this method allows for the separation of all cellular proteins on a single large gel. A major advantage of this method is that it often distinguishes between different isoforms of a particular protein - e.g. a protein that has been phosphorylated (by addition of a negatively charged group). Proteins that have been separated can be cut out of the gel and then analysed by mass spectrometry, which identifies the protein.

Please refer to reference articles for examples of the application of 2-D SDS PAGE.

Medical diagnostic applications

- The confirmatory HIV test employs a Western blot to detect anti-HIV antibody in a human serum sample. Proteins from known HIV-infected cells are separated and blotted on a membrane as above. Then, the serum to be tested is applied in the primary antibody incubation step; free antibody is washed away, and a secondary anti-human antibody linked to an enzyme signal is added. The stained bands then indicate the proteins to which the patient's serum contains antibody.
- A Western blot is also used as the definitive test for Bovine spongiform encephalopathy (BSE, commonly referred to as 'mad cow disease').
- Some forms of Lyme disease testing employ Western blotting.
- Western blot can also be used as a confirmatory test for Hepatitis B infection.
- In veterinary medicine, Western blot is sometimes used to confirm FIV+ status in cats.

Protocols

- Western Blotting Protocol from Protocolmonkey
- Modified Western blotting protocols from Biotechniques

- Western Immunoblot protocols from Cell Signaling Technology
- Dr. Mark Barton Frank Lab protocol
- Kamps's Western Blotting Protocol
- western blot protocols, troubleshooting, scientific resources and research articles

Southern blot

A **Southern blot** is a method routinely used in molecular biology for detection of a specific DNA sequence in DNA samples. Southern blotting combines transfer of electrophoresis-separated DNA fragments to a filter membrane and subsequent fragment detection by probe hybridization. The method is named after its inventor, the British biologist Edwin Southern. Other blotting methods (i.e., Western blot, Northern blot, Eastern blot, Southwestern blot) that employ similar principles, but using RNA or protein, have later been named in reference to Edwin Southern's name. As the technique was eponymously named, Southern blot is capitalized as is conventional for proper nouns. The names for other blotting methods may follow this convention, by analogy.

Method

1. Restriction endonucleases are used to cut high-molecular-weight DNA strands into smaller fragments.
2. The DNA fragments are then electrophoresed on an agarose gel to separate them by size.
3. If some of the DNA fragments are larger than 15 kb, then prior to blotting, the gel may be treated with an acid, such as dilute HCl, which depurinates the DNA fragments, breaking the DNA into smaller pieces, thus allowing more efficient transfer from the gel to membrane.
4. If alkaline transfer methods are used, the DNA gel is placed into an alkaline solution (typically containing sodium hydroxide) to denature the double-stranded DNA. The denaturation in an alkaline environment may improve binding of the negatively charged DNA to a positively charged membrane, separating it into single DNA strands for later hybridization to the probe, and destroys any residual RNA that may still be present in the DNA. The choice of alkaline over neutral transfer methods, however, is often empirical and may result in equivalent results.
5. A sheet of nitrocellulose (or, alternatively, nylon) membrane is placed on top of (or below, depending on the direction of the transfer) the gel. Pressure is applied evenly to the gel (either using suction, or by placing a stack of paper towels and a weight on top of the membrane and gel), to ensure good and even contact between gel and membrane. If transferring by suction 20X SSC buffer is used to ensure a seal and prevent drying of the gel. Buffer transfer by capillary action from a region of high water potential to a region of low water potential (usually filter paper and paper tissues) is then used to move the DNA from the gel on to the

membrane; ion exchange interactions bind the DNA to the membrane due to the negative charge of the DNA and positive charge of the membrane.

6. The membrane is then baked in a vacuum or regular oven at 80 °C for 2 hours (standard conditions; nitrocellulose or nylon membrane) or exposed to ultraviolet radiation (nylon membrane) to permanently attach the transferred DNA to the membrane.
7. The membrane is then exposed to a hybridization probe—a single DNA fragment with a specific sequence whose presence in the target DNA is to be determined. The probe DNA is labelled so that it can be detected, usually by incorporating radioactivity or tagging the molecule with a fluorescent or chromogenic dye. In some cases, the hybridization probe may be made from RNA, rather than DNA. To ensure the specificity of the binding of the probe to the sample DNA, most common hybridization methods use salmon or herring sperm DNA for blocking of the membrane surface and target DNA, deionized formamide, and detergents such as SDS to reduce non-specific binding of the probe.
8. After hybridization, excess probe is washed from the membrane (typically using SSC buffer), and the pattern of hybridization is visualized on X-ray film by autoradiography in the case of a radioactive or fluorescent probe, or by development of color on the membrane if a chromogenic detection method is used.

Result

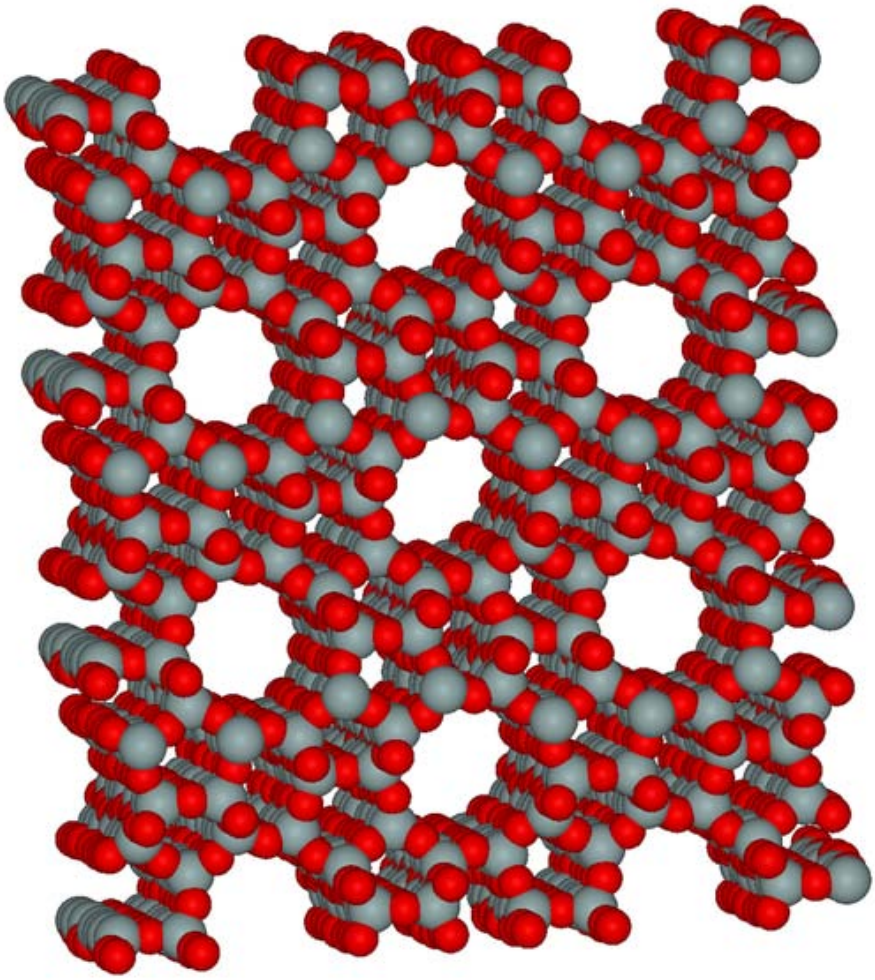
Hybridization of the probe to a specific DNA fragment on the filter membrane indicates that this fragment contains DNA sequence that is complementary to the probe.

The transfer step of the DNA from the electrophoresis gel to a membrane permits easy binding of the labeled hybridization probe to the size-fractionated DNA. It also allows for the fixation of the target-probe hybrids, required for analysis by autoradiography or other detection methods.

Southern blots performed with restriction enzyme-digested genomic DNA may be used to determine the number of sequences (e.g., gene copies) in a genome. A probe that hybridizes only to a single DNA segment that has not been cut by the restriction enzyme will produce a single band on a Southern blot, whereas multiple bands will likely be observed when the probe hybridizes to several highly similar sequences (e.g., those that may be the result of sequence duplication). Modification of the hybridization conditions (for example, increasing the hybridization temperature or decreasing salt concentration) may be used to increase specificity and decrease hybridization of the probe to sequences that are less than 100% similar.

Chapter- 8

X-ray Crystallography



X-ray crystallography can locate every atom in a zeolite, an aluminosilicate with many important applications, such as water purification.

X-ray crystallography is a method of determining the arrangement of atoms within a crystal, in which a beam of X-rays strikes a crystal and diffracts into many specific directions. From the angles and intensities of these diffracted beams, a crystallographer can produce a three-dimensional picture of the density of electrons within the crystal. From this electron density, the mean positions of the atoms in the crystal can be determined, as well as their chemical bonds, their disorder and various other information.

Since many materials can form crystals — such as salts, metals, minerals, semiconductors, as well as various inorganic, organic and biological molecules — X-ray crystallography has been fundamental in the development of many scientific fields. In its first decades of use, this method determined the size of atoms, the lengths and types of chemical bonds, and the atomic-scale differences among various materials, especially minerals and alloys. The method also revealed the structure and functioning of many biological molecules, including vitamins, drugs, proteins and nucleic acids such as DNA. X-ray crystallography is still the chief method for characterizing the atomic structure of new materials and in discerning materials that appear similar by other experiments. X-ray crystal structures can also account for unusual electronic or elastic properties of a material, shed light on chemical interactions and processes, or serve as the basis for designing pharmaceuticals against diseases.

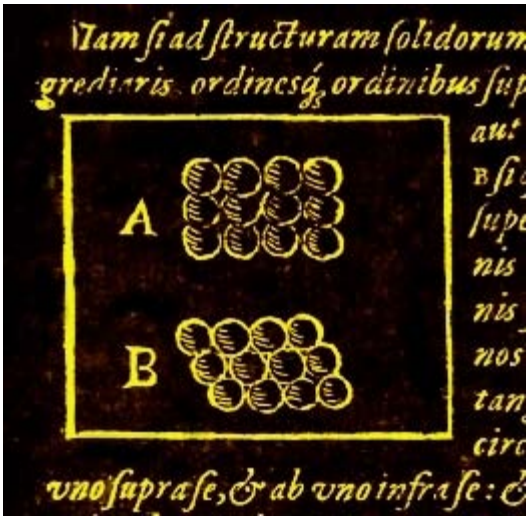
In an X-ray diffraction measurement, a crystal is mounted on a goniometer and gradually rotated while being bombarded with X-rays, producing a diffraction pattern of regularly spaced spots known as *reflections*. The two-dimensional images taken at different rotations are converted into a three-dimensional model of the density of electrons within the crystal using the mathematical method of Fourier transforms, combined with chemical data known for the sample. Poor resolution (fuzziness) or even errors may result if the crystals are too small, or not uniform enough in their internal makeup.

X-ray crystallography is related to several other methods for determining atomic structures. Similar diffraction patterns can be produced by scattering electrons or neutrons, which are likewise interpreted as a Fourier transform. If single crystals of sufficient size cannot be obtained, various other X-ray methods can be applied to obtain less detailed information; such methods include fiber diffraction, powder diffraction and small-angle X-ray scattering (SAXS). If the material under investigation is only available in the form of nanocrystalline powders or suffers from poor crystallinity, the methods of electron crystallography can be applied for determining the atomic structure.

For all above mentioned X-ray diffraction methods, the scattering is elastic; the scattered X-rays have the same wavelength as the incoming X-ray. By contrast, *inelastic* X-ray scattering methods are useful in studying excitations of the sample, rather than the distribution of its atoms.

History

Early scientific history of crystals and X-rays

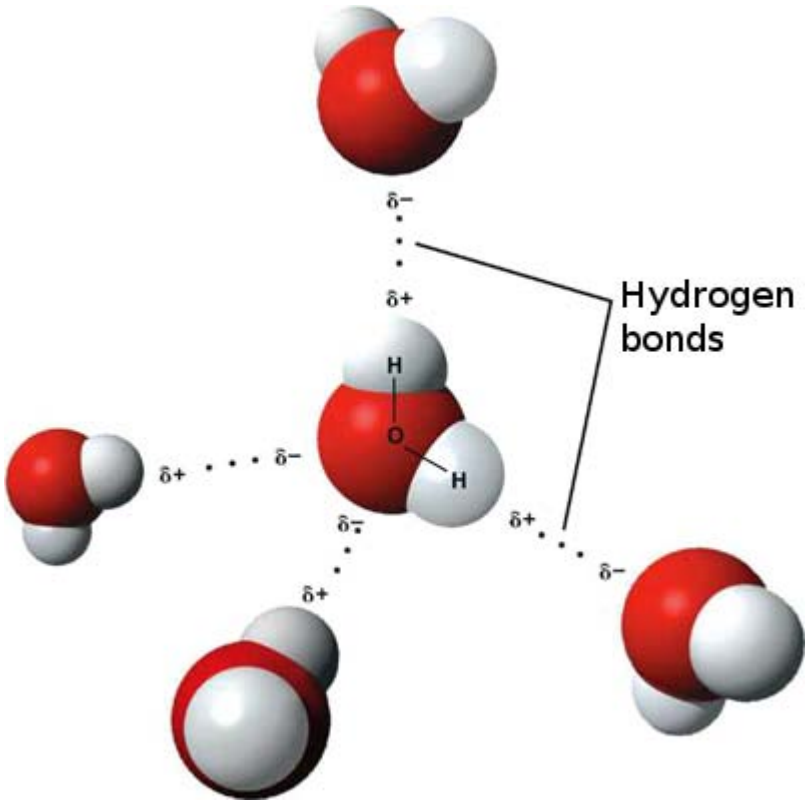


Drawing of square (Figure A, above) and hexagonal (Figure B, below) packing from Kepler's work, *Strena seu de Nive Sexangula*.

Crystals have long been admired for their regularity and symmetry, but they were not investigated scientifically until the 17th century. Johannes Kepler hypothesized in his work *Strena seu de Nive Sexangula* (1611) that the hexagonal symmetry of snowflake crystals was due to a regular packing of spherical water particles.

As shown by X-ray crystallography, the hexagonal symmetry of snowflakes results from the tetrahedral arrangement of hydrogen bonds about each water molecule. The water molecules are arranged similarly to the silicon atoms in the tridymite polymorph of SiO_2. The resulting crystal structure has hexagonal symmetry when viewed along a principal axis.

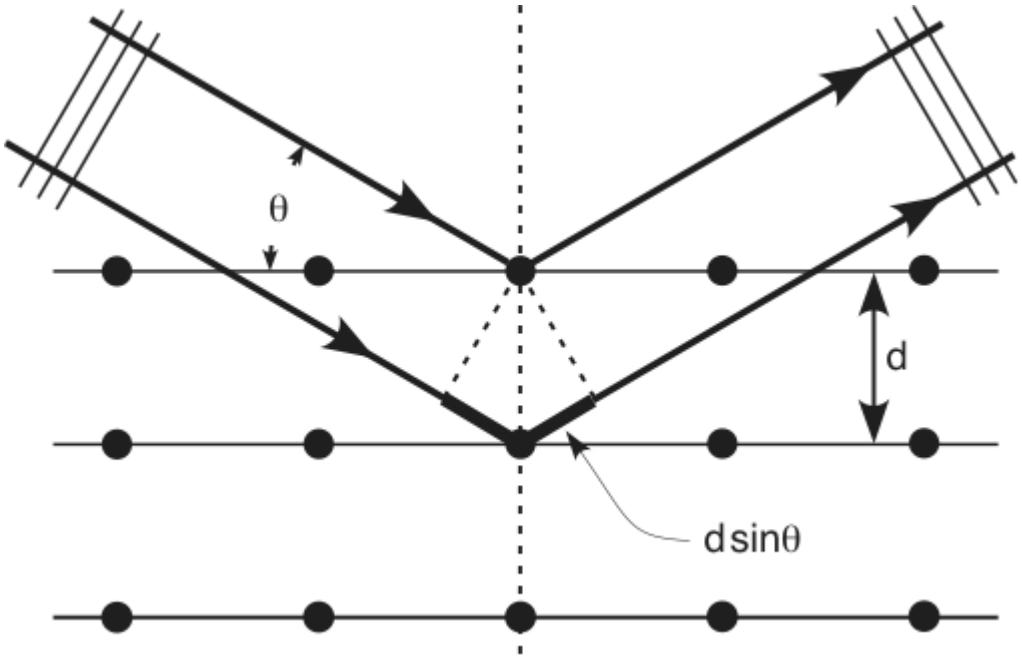
Crystal symmetry was first investigated experimentally by Nicolas Steno (1669), who showed that the angles between the faces are the same in every exemplar of a particular type of crystal, and by René Just Haüy (1784), who discovered that every face of a crystal can be described by simple stacking patterns of blocks of the same shape and size. Hence, William Hallowes Miller in 1839 was able to give each face a unique label of three small integers, the Miller indices which are still used today for identifying crystal faces. Haüy's study led to the correct idea that crystals are a regular three-dimensional array (a Bravais lattice) of atoms and molecules; a single unit cell is repeated indefinitely along three principal directions that are not necessarily perpendicular. In the 19th century, a complete catalog of the possible symmetries of a crystal was worked out by Johann Hessel, Auguste Bravais, Yevgraf Fyodorov, Arthur Schönflies and (belatedly) William Barlow. From the available data and physical reasoning, Barlow proposed several crystal structures in the 1880s that were validated later by X-ray crystallography; however, the available data were too scarce in the 1880s to accept his models as conclusive.



X-ray crystallography shows the arrangement of water molecules in ice, revealing the hydrogen bonds that hold the solid together. Few other methods can determine the structure of matter with such sub-atomic precision (*resolution*).

X-rays were discovered by Wilhelm Conrad Röntgen in 1895, just as the studies of crystal symmetry were being concluded. Physicists were initially uncertain of the nature of X-rays, although it was soon suspected (correctly) that they were waves of electromagnetic radiation, in other words, another form of light. At that time, the wave model of light — specifically, the Maxwell theory of electromagnetic radiation — was well accepted among scientists, and experiments by Charles Glover Barkla showed that X-rays exhibited phenomena associated with electromagnetic waves, including transverse polarization and spectral lines akin to those observed in the visible wavelengths. Single-slit experiments in the laboratory of Arnold Sommerfeld suggested the wavelength of X-rays was about 1 Angström. However, X-rays are composed of photons, and thus are not only waves of electromagnetic radiation but also exhibit particle-like properties. The photon concept was introduced by Albert Einstein in 1905, but it was not broadly accepted until 1922, when Arthur Compton confirmed it by the scattering of X-rays from electrons. Therefore, these particle-like properties of X-rays, such as their ionization of gases, caused William Henry Bragg to argue in 1907 that X-rays were *not* electromagnetic radiation. Nevertheless, Bragg's view was not broadly accepted and the observation of X-ray diffraction in 1912 confirmed for most scientists that X-rays were a form of electromagnetic radiation.

X-ray analysis of crystals



The incoming beam (coming from upper left) causes each scatterer to re-radiate a small portion of its intensity as a spherical wave. If scatterers are arranged symmetrically with a separation *d*, these spherical waves will be in sync (add constructively) only in directions where their path-length difference 2*d* sin θ equals an integer multiple of the wavelength λ. In that case, part of the incoming beam is deflected by an angle 2θ, producing a *reflection* spot in the diffraction pattern.

Crystals are regular arrays of atoms, and X-rays can be considered waves of electromagnetic radiation. Atoms scatter X-ray waves, primarily through the atoms' electrons. Just as an ocean wave striking a lighthouse produces secondary circular waves emanating from the lighthouse, so an X-ray striking an electron produces secondary spherical waves emanating from the electron. This phenomenon is known as elastic scattering, and the electron (or lighthouse) is known as the *scatterer*. A regular array of scatterers produces a regular array of spherical waves. Although these waves cancel one another out in most directions through destructive interference, they add constructively in a few specific directions, determined by Bragg's law:

$$2d \sin \theta = n\lambda$$

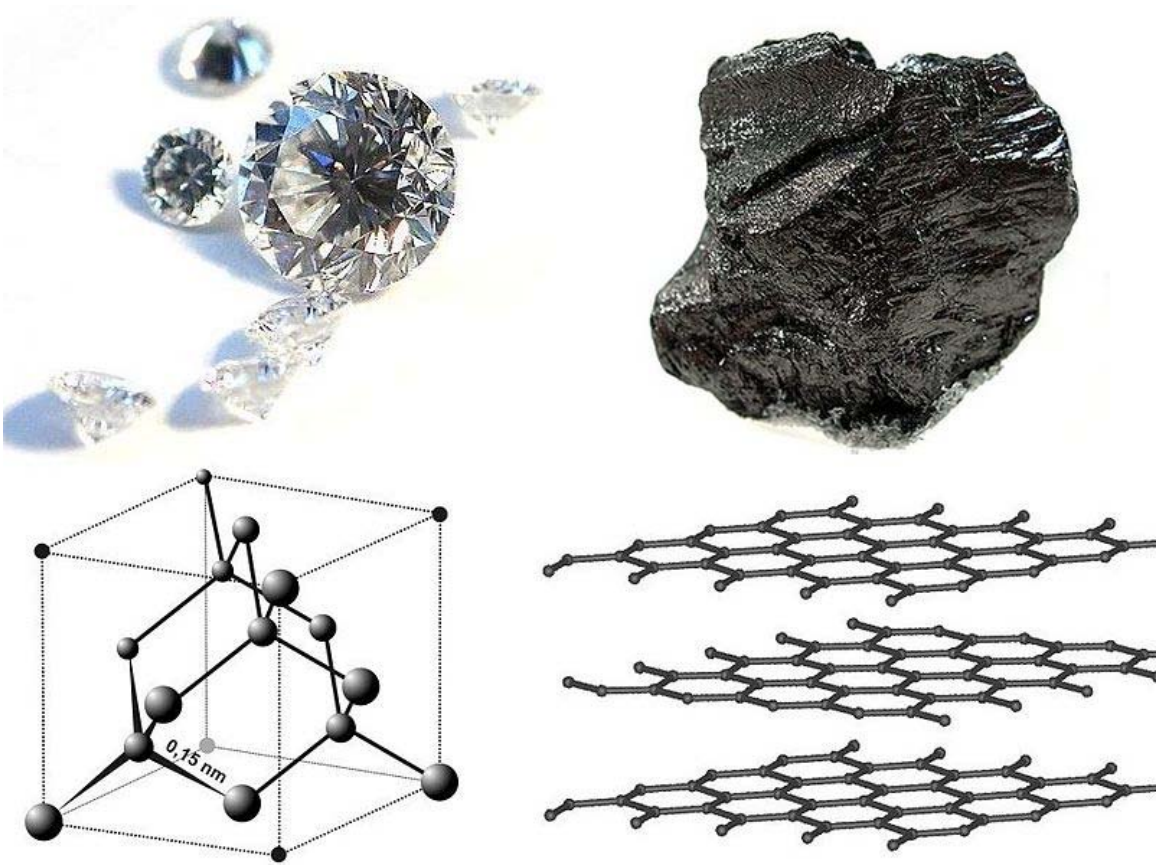
Here *d* is the spacing between diffracting planes, θ is the incident angle, *n* is any integer, and λ is the wavelength of the beam. These specific directions appear as spots on the diffraction pattern called *reflections*. Thus, X-ray diffraction results from an electromagnetic wave (the X-ray) impinging on a regular array of scatterers (the repeating arrangement of atoms within the crystal).

X-rays are used to produce the diffraction pattern because their wavelength λ is typically the same order of magnitude (1-100 Ångströms) as the spacing *d* between planes in the crystal. In principle, any wave impinging on a regular array of scatterers produces diffraction, as predicted first by Francesco Maria Grimaldi in 1665. To produce significant diffraction, the spacing between the scatterers and the wavelength of the impinging wave should be similar in size. For illustration, the diffraction of sunlight through a bird's feather was first reported by James Gregory in the later 17th century. The first artificial diffraction gratings for visible light were constructed by David Rittenhouse in 1787, and Joseph von Fraunhofer in 1821. However, visible light has too long a wavelength (typically, 5500 Ångströms) to observe diffraction from crystals. Prior to the first X-ray diffraction experiments, the spacings between lattice planes in a crystal were not known with certainty.

The idea that crystals could be used as a diffraction grating for X-rays arose in 1912 in a conversation between Paul Peter Ewald and Max von Laue in the English Garden in Munich. Ewald had proposed a resonator model of crystals for his thesis, but this model could not be validated using visible light, since the wavelength was much larger than the spacing between the resonators. Von Laue realized that electromagnetic radiation of a shorter wavelength was needed to observe such small spacings, and suggested that X-rays might have a wavelength comparable to the unit-cell spacing in crystals. Von Laue worked with two technicians, Walter Friedrich and his assistant Paul Knipping, to shine a beam of X-rays through a copper sulfate crystal and record its diffraction on a photographic plate. After being developed, the plate showed a large number of well-defined spots arranged in a pattern of intersecting circles around the spot produced by the central beam. Von Laue developed a law that connects the scattering angles and the size and orientation of the unit-cell spacings in the crystal, for which he was awarded the Nobel Prize in Physics in 1914.

As described in the mathematical derivation below, the X-ray scattering is determined by the density of electrons within the crystal. Since the energy of an X-ray is much greater than that of a valence electron, the scattering may be modeled as Thomson scattering, the interaction of an electromagnetic ray with a free electron. This model is generally adopted to describe the polarization of the scattered radiation. The intensity of Thomson scattering declines as $1/m^2$ with the mass m of the charged particle that is scattering the radiation; hence, the atomic nuclei, which are thousands of times heavier than an electron, contribute negligibly to the scattered X-rays.

Development from 1912 to 1920



Although diamonds (top left) and graphite (top right) are identical in chemical composition — being both pure carbon — X-ray crystallography revealed the arrangement of their atoms (bottom) accounts for their different properties. In diamond, the carbon atoms are arranged tetrahedrally and held together by single covalent bonds, making it strong in all directions. By contrast, graphite is composed of stacked sheets. Within the sheet, the bonding is covalent and has hexagonal symmetry, but there are no covalent bonds between the sheets, making graphite easy to cleave into flakes.

After Von Laue's pioneering research, the field developed rapidly, most notably by physicists William Lawrence Bragg and his father William Henry Bragg. In 1912-1913, the younger Bragg developed Bragg's law, which connects the observed scattering with

reflections from evenly spaced planes within the crystal. The Braggs, father and son, shared the 1915 Nobel Prize in Physics for their work in crystallography. The earliest structures were generally simple and marked by one-dimensional symmetry. However, as computational and experimental methods improved over the next decades, it became feasible to deduce reliable atomic positions for more complicated two- and three-dimensional arrangements of atoms in the unit-cell.

The potential of X-ray crystallography for determining the structure of molecules and minerals — then only known vaguely from chemical and hydrodynamic experiments — was realized immediately. The earliest structures were simple inorganic crystals and minerals, but even these revealed fundamental laws of physics and chemistry. The first atomic-resolution structure to be "solved" (i.e. determined) in 1914 was that of table salt. The distribution of electrons in the table-salt structure showed that crystals are not necessarily composed of covalently bonded molecules, and proved the existence of ionic compounds. The structure of diamond was solved in the same year, proving the tetrahedral arrangement of its chemical bonds and showing that the length of C–C single bond was 1.52 Ångströms. Other early structures included copper, calcium fluoride (CaF_2, also known as *fluorite*), calcite ($CaCO_3$) and pyrite (FeS_2) in 1914; spinel ($MgAl_2O_4$) in 1915; the rutile and anatase forms of titanium dioxide (TiO_2) in 1916; pyrochroite $Mn(OH)_2$ and, by extension, brucite $Mg(OH)_2$ in 1919;. Also in 1919 sodium nitrate ($NaNO_3$) and cesium dichloroiodide ($CsICl_2$) were determined by Ralph Walter Graystone Wyckoff, and the wurtzite (hexagonal ZnS) structure became known in 1920.

The structure of graphite was solved in 1916 by the related method of powder diffraction, which was developed by Peter Debye and Paul Scherrer and, independently, by Albert Hull in 1917. The structure of graphite was determined from single-crystal diffraction in 1924 by two groups independently. Hull also used the powder method to determine the structures of various metals, such as iron and magnesium.

Contributions to chemistry and material science

X-ray crystallography has led to a better understanding of chemical bonds and non-covalent interactions. The initial studies revealed the typical radii of atoms, and confirmed many theoretical models of chemical bonding, such as the tetrahedral bonding of carbon in the diamond structure, the octahedral bonding of metals observed in ammonium hexachloroplatinate (IV), and the resonance observed in the planar carbonate group and in aromatic molecules. Kathleen Lonsdale's 1928 structure of hexamethylbenzene established the hexagonal symmetry of benzene and showed a clear difference in bond length between the aliphatic C–C bonds and aromatic C–C bonds; this finding led to the idea of resonance between chemical bonds, which had profound consequences for the development of chemistry. Her conclusions were anticipated by William Henry Bragg, who published models of naphthalene and anthracene in 1921 based on other molecules, an early form of molecular replacement.

Also in the 1920s, Victor Moritz Goldschmidt and later Linus Pauling developed rules for eliminating chemically unlikely structures and for determining the relative sizes of atoms.

These rules led to the structure of brookite (1928) and an understanding of the relative stability of the rutile, brookite and anatase forms of titanium dioxide.

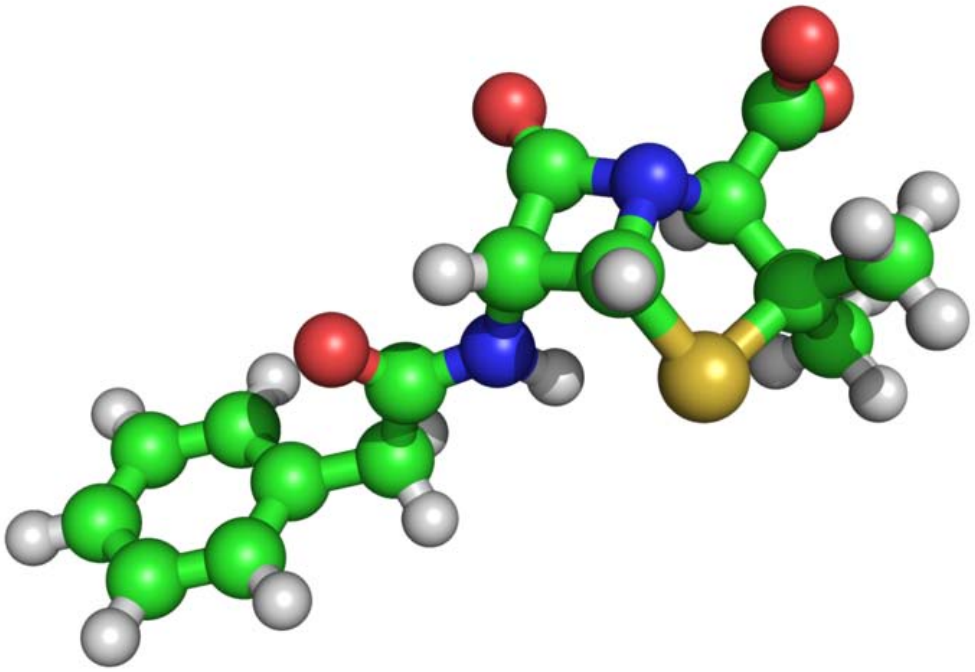
The distance between two bonded atoms is a sensitive measure of the bond strength and its bond order; thus, X-ray crystallographic studies have led to the discovery of even more exotic types of bonding in inorganic chemistry, such as metal-metal double bonds, metal-metal quadruple bonds, and three-center, two-electron bonds. X-ray crystallography — or, strictly speaking, an inelastic Compton scattering experiment — has also provided evidence for the partly covalent character of hydrogen bonds. In the field of organometallic chemistry, the X-ray structure of ferrocene initiated scientific studies of sandwich compounds, while that of Zeise's salt stimulated research into "back bonding" and metal-pi complexes. Finally, X-ray crystallography had a pioneering role in the development of supramolecular chemistry, particularly in clarifying the structures of the crown ethers and the principles of host-guest chemistry.

In material sciences, many complicated inorganic and organometallic systems have been analyzed using single-crystal methods, such as fullerenes, metalloporphyrins, and other complicated compounds. Single-crystal diffraction is also used in the pharmaceutical industry, due to recent problems with polymorphs. The major factors affecting the quality of single-crystal structures are the crystal's size and regularity; recrystallization is a commonly used technique to improve these factors in small-molecule crystals. The Cambridge Structural Database contains over 500,000 structures; over 99% of these structures were determined by X-ray diffraction.

Mineralogy and metallurgy

Since the 1920s, X-ray diffraction has been the principal method for determining the arrangement of atoms in minerals and metals. The application of X-ray crystallography to mineralogy began with the structure of garnet, which was determined in 1924 by Menzer. A systematic X-ray crystallographic study of the silicates was undertaken in the 1920s. This study showed that, as the Si/O ratio is altered, the silicate crystals exhibit significant changes in their atomic arrangements. Machatschki extended these insights to minerals in which aluminium substitutes for the silicon atoms of the silicates. The first application of X-ray crystallography to metallurgy likewise occurred in the mid-1920s. Most notably, Linus Pauling's structure of the alloy Mg_2Sn led to his theory of the stability and structure of complex ionic crystals.

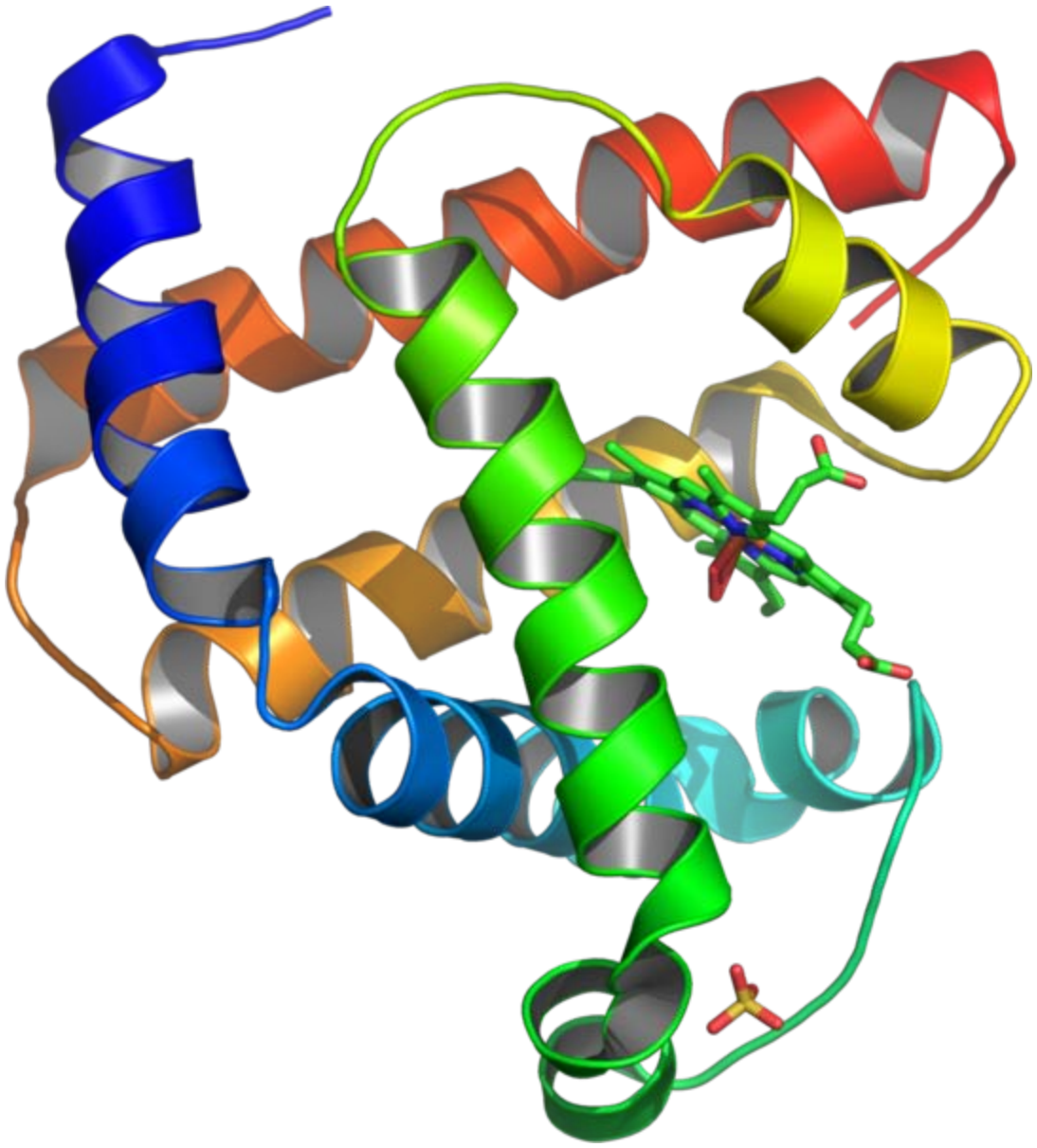
Early organic and small biological molecules



The three-dimensional structure of penicillin, for which Dorothy Crowfoot Hodgkin was awarded the Nobel Prize in Chemistry in 1964. The green, white, red, yellow and blue spheres represent atoms of carbon, hydrogen, oxygen, sulfur and nitrogen, respectively.

The first structure of an organic compound, hexamethylenetetramine, was solved in 1923. This was followed by several studies of long-chain fatty acids, which are an important component of biological membranes. In the 1930s, the structures of much larger molecules with two-dimensional complexity began to be solved. A significant advance was the structure of phthalocyanine, a large planar molecule that is closely related to porphyrin molecules important in biology, such as heme, corrin and chlorophyll.

X-ray crystallography of biological molecules took off with Dorothy Crowfoot Hodgkin, who solved the structures of cholesterol (1937), vitamin B12 (1945) and penicillin (1954), for which she was awarded the Nobel Prize in Chemistry in 1964. In 1969, she succeeded in solving the structure of insulin, on which she worked for over thirty years.



Ribbon diagram of the structure of myoglobin, showing colored alpha helices. Such proteins are long, linear molecules with thousands of atoms; yet the relative position of each atom has been determined with sub-atomic resolution by X-ray crystallography. Since it is difficult to visualize all the atoms at once, the ribbon shows the rough path of the protein polymer from its N-terminus (blue) to its C-terminus (red).

Biological macromolecular crystallography

Crystal structures of proteins (which are irregular and hundreds of times larger than cholesterol) began to be solved in the late 1950s, beginning with the structure of sperm whale myoglobin by Max Perutz and Sir John Cowdery Kendrew, for which they were awarded the Nobel Prize in Chemistry in 1962. Since that success, over 48970 X-ray crystal structures of proteins, nucleic acids and other biological molecules have been

determined. For comparison, the nearest competing method in terms of structures analyzed is nuclear magnetic resonance (NMR) spectroscopy, which has resolved 7806 chemical structures. Moreover, crystallography can solve structures of arbitrarily large molecules, whereas solution-state NMR is restricted to relatively small ones (less than 70 kDa). X-ray crystallography is now used routinely by scientists to determine how a pharmaceutical drug interacts with its protein target and what changes might improve it. However, intrinsic membrane proteins remain challenging to crystallize because they require detergents or other means to solubilize them in isolation, and such detergents often interfere with crystallization. Such membrane proteins are a large component of the genome and include many proteins of great physiological importance, such as ion channels and receptors.

Relationship to other scattering techniques

Elastic vs. inelastic scattering

X-ray crystallography is a form of elastic scattering; the outgoing X-rays have the same energy, and thus same wavelength, as the incoming X-rays, only with altered direction. By contrast, *inelastic scattering* occurs when energy is transferred from the incoming X-ray to the crystal, e.g., by exciting an inner-shell electron to a higher energy level. Such inelastic scattering reduces the energy (or increases the wavelength) of the outgoing beam. Inelastic scattering is useful for probing such excitations of matter, but not in determining the distribution of scatterers within the matter, which is the goal of X-ray crystallography.

X-rays range in wavelength from 10 to 0.01 nanometers; a typical wavelength used for crystallography is 1 Å (0.1 nm), which is on the scale of covalent chemical bonds and the radius of a single atom. Longer-wavelength photons (such as ultraviolet radiation) would not have sufficient resolution to determine the atomic positions. At the other extreme, shorter-wavelength photons such as gamma rays are difficult to produce in large numbers, difficult to focus, and interact too strongly with matter, producing particle-antiparticle pairs. Therefore, X-rays are the "sweetspot" for wavelength when determining atomic-resolution structures from the scattering of electromagnetic radiation.

Other X-ray techniques

Other forms of elastic X-ray scattering include powder diffraction, SAXS and several types of X-ray fiber diffraction, which was used by Rosalind Franklin in determining the double-helix structure of DNA. In general, single-crystal X-ray diffraction offers more structural information than these other techniques; however, it requires a sufficiently large and regular crystal, which is not always available.

These scattering methods generally use *monochromatic* X-rays, which are restricted to a single wavelength with minor deviations. A broad spectrum of X-rays (that is, a blend of X-rays with different wavelengths) can also be used to carry out X-ray diffraction, a technique known as the Laue method. This is the method used in the original discovery of

X-ray diffraction. Laue scattering provides much structural information with only a short exposure to the X-ray beam, and is therefore used in structural studies of very rapid events (Time resolved crystallography). However, it is not as well-suited as monochromatic scattering for determining the full atomic structure of a crystal and therefore works better with crystals with relatively simple atomic arrangements.

The Laue back reflection mode records X-rays scattered backwards from a broad spectrum source. This is useful if the sample is too thick for X-rays to transmit through it. The diffracting planes in the crystal are determined by knowing that the normal to the diffracting plane bisects the angle between the incident beam and the diffracted beam. A Greninger chart can be used to interpret the back reflection Laue photograph.

Electron and neutron diffraction

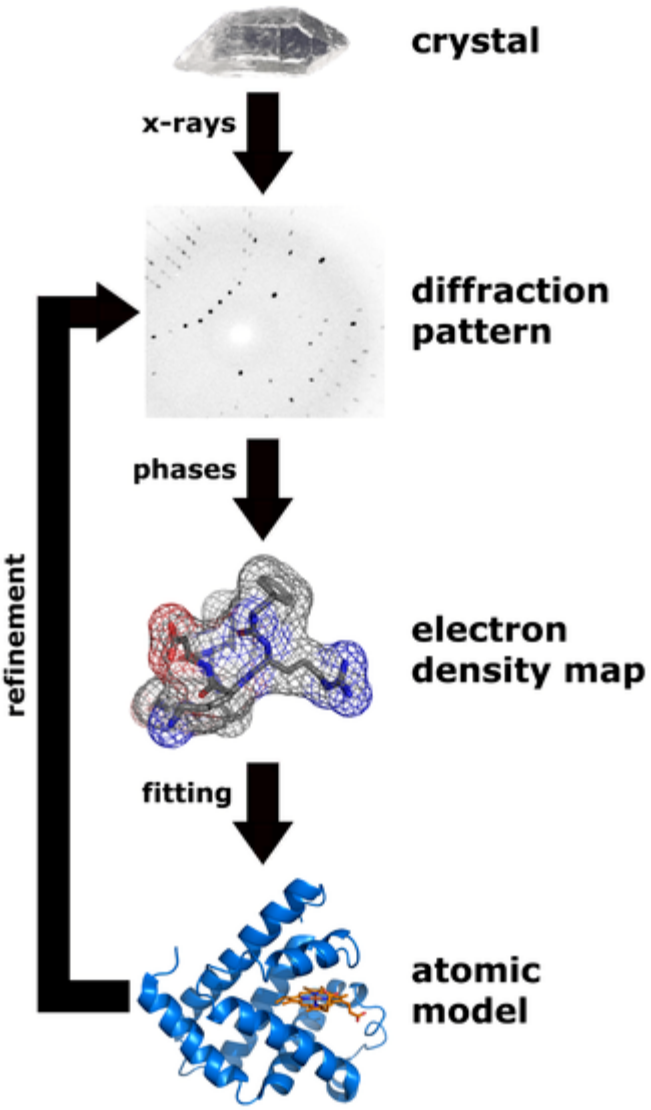
Other particles, such as electrons and neutrons, may be used to produce a diffraction pattern. Although electron, neutron, and X-ray scattering are based on different physical processes, the resulting diffraction patterns are analyzed using the same coherent diffraction imaging techniques.

As derived below, the electron density within the crystal and the diffraction patterns are related by a simple mathematical method, the Fourier transform, which allows the density to be calculated relatively easily from the patterns. However, this works only if the scattering is *weak*, i.e., if the scattered beams are much less intense than the incoming beam. Weakly scattered beams pass through the remainder of the crystal without undergoing a second scattering event. Such re-scattered waves are called "secondary scattering" and hinder the analysis. Any sufficiently thick crystal will produce secondary scattering, but since X-rays interact relatively weakly with the electrons, this is generally not a significant concern. By contrast, electron beams may produce strong secondary scattering even for relatively thin crystals (>100 nm). Since this thickness corresponds to the diameter of many viruses, a promising direction is the electron diffraction of isolated macromolecular assemblies, such as viral capsids and molecular machines, which may be carried out with a cryo-electron microscope. Moreover the strong interaction of electrons with matter (about 1000 times stronger than for X-rays) allows determination of the atomic structure of extremely small volumes. The field of applications for electron crystallography ranges from bio molecules like membrane proteins over organic thin films to the complex structures of (nanocrystalline) intermetallic compounds and zeolites.

Neutron diffraction is an excellent method for structure determination, although it has been difficult to obtain intense, monochromatic beams of neutrons in sufficient quantities. Traditionally, nuclear reactors have been used, although the new Spallation Neutron Source holds much promise in the near future. Being uncharged, neutrons scatter much more readily from the atomic nuclei rather than from the electrons. Therefore, neutron scattering is very useful for observing the positions of light atoms with few electrons, especially hydrogen, which is essentially invisible in the X-ray diffraction. Neutron scattering also has the remarkable property that the solvent can be made invisible by adjusting the ratio of normal water, H_2O, and heavy water, D_2O.

Methods

Overview of single-crystal X-ray diffraction



Workflow for solving the structure of a molecule by X-ray crystallography.

The oldest and most precise method of X-ray crystallography is *single-crystal X-ray diffraction*, in which a beam of X-rays strikes a single crystal, producing scattered beams. When they land on a piece of film or other detector, these beams make a *diffraction pattern* of spots; the strengths and angles of these beams are recorded as the crystal is gradually rotated. Each spot is called a *reflection*, since it corresponds to the reflection of the X-rays from one set of evenly spaced planes within the crystal. For single crystals of

sufficient purity and regularity, X-ray diffraction data can determine the mean chemical bond lengths and angles to within a few thousandths of an Ångström and to within a few tenths of a degree, respectively. The atoms in a crystal are not static, but oscillate about their mean positions, usually by less than a few tenths of an Ångström. X-ray crystallography allows measuring the size of these oscillations.

Procedure

The technique of single-crystal X-ray crystallography has three basic steps. The first — and often most difficult — step is to obtain an adequate crystal of the material under study. The crystal should be sufficiently large (typically larger than 0.1 mm in all dimensions), pure in composition and regular in structure, with no significant internal imperfections such as cracks or twinning.

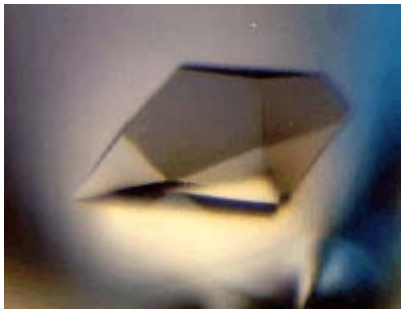
In the second step, the crystal is placed in an intense beam of X-rays, usually of a single wavelength (*monochromatic X-rays*), producing the regular pattern of reflections. As the crystal is gradually rotated, previous reflections disappear and new ones appear; the intensity of every spot is recorded at every orientation of the crystal. Multiple data sets may have to be collected, with each set covering slightly more than half a full rotation of the crystal and typically containing tens of thousands of reflections.

In the third step, these data are combined computationally with complementary chemical information to produce and refine a model of the arrangement of atoms within the crystal. The final, refined model of the atomic arrangement — now called a *crystal structure* — is usually stored in a public database.

Limitations

As the crystal's repeating unit, its unit cell, becomes larger and more complex, the atomic-level picture provided by X-ray crystallography becomes less well-resolved (more "fuzzy") for a given number of observed reflections. Two limiting cases of X-ray crystallography—"small-molecule" and "macromolecular" crystallography—are often discerned. *Small-molecule crystallography* typically involves crystals with fewer than 100 atoms in their asymmetric unit; such crystal structures are usually so well resolved that the atoms can be discerned as isolated "blobs" of electron density. By contrast, *macromolecular crystallography* often involves tens of thousands of atoms in the unit cell. Such crystal structures are generally less well-resolved (more "smeared out"); the atoms and chemical bonds appear as tubes of electron density, rather than as isolated atoms. In general, small molecules are also easier to crystallize than macromolecules; however, X-ray crystallography has proven possible even for viruses with hundreds of thousands of atoms.

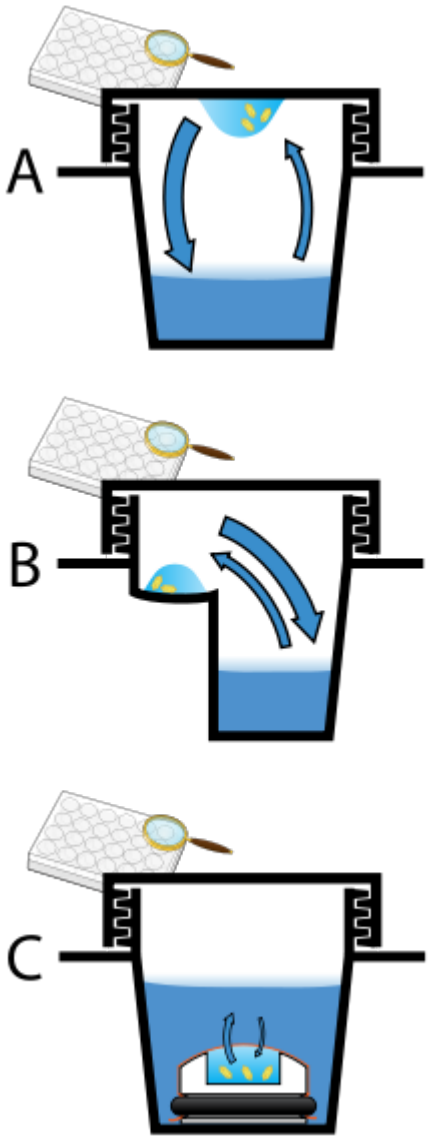
Crystallization



A protein crystal seen under a microscope. Crystals used in X-ray crystallography may be smaller than a millimeter across.

Although crystallography can be used to characterize the disorder in an impure or irregular crystal, crystallography generally requires a pure crystal of high regularity to solve the structure of a complicated arrangement of atoms. Pure, regular crystals can sometimes be obtained from natural or synthetic materials, such as samples of metals, minerals or other macroscopic materials. The regularity of such crystals can sometimes be improved with macromolecular crystal annealing and other methods. However, in many cases, obtaining a diffraction-quality crystal is the chief barrier to solving its atomic-resolution structure.

Small-molecule and macromolecular crystallography differ in the range of possible techniques used to produce diffraction-quality crystals. Small molecules generally have few degrees of conformational freedom, and may be crystallized by a wide range of methods, such as chemical vapor deposition and recrystallization. By contrast, macromolecules generally have many degrees of freedom and their crystallization must be carried out to maintain a stable structure. For example, proteins and larger RNA molecules cannot be crystallized if their tertiary structure has been unfolded; therefore, the range of crystallization conditions is restricted to solution conditions in which such molecules remain folded.



Three methods of preparing crystals, A: Hanging drop. B: Sitting drop. C: Microdialysis

Protein crystals are almost always grown in solution. The most common approach is to lower the solubility of its component molecules very gradually; if this is done too quickly, the molecules will precipitate from solution, forming a useless dust or amorphous gel on the bottom of the container. Crystal growth in solution is characterized by two steps: *nucleation* of a microscopic crystallite (possibly having only 100 molecules), followed by *growth* of that crystallite, ideally to a diffraction-quality crystal. The solution conditions that favor the first step (nucleation) are not always the same conditions that favor the second step (subsequent growth). The crystallographer's goal is to identify solution conditions that favor the development of a single, large crystal, since larger crystals offer improved resolution of the molecule. Consequently, the solution conditions should *disfavor* the first step (nucleation) but *favor* the second (growth), so

that only one large crystal forms per droplet. If nucleation is favored too much, a shower of small crystallites will form in the droplet, rather than one large crystal; if favored too little, no crystal will form whatsoever.

It is extremely difficult to predict good conditions for nucleation or growth of well-ordered crystals. In practice, favorable conditions are identified by *screening*; a very large batch of the molecules is prepared, and a wide variety of crystallization solutions are tested. Hundreds, even thousands, of solution conditions are generally tried before finding the successful one. The various conditions can use one or more physical mechanisms to lower the solubility of the molecule; for example, some may change the pH, some contain salts of the Hofmeister series or chemicals that lower the dielectric constant of the solution, and still others contain large polymers such as polyethylene glycol that drive the molecule out of solution by entropic effects. It is also common to try several temperatures for encouraging crystallization, or to gradually lower the temperature so that the solution becomes supersaturated. These methods require large amounts of the target molecule, as they use high concentration of the molecule(s) to be crystallized. Due to the difficulty in obtaining such large quantities (milligrams) of crystallization grade protein, robots have been developed that are capable of accurately dispensing crystallization trial drops that are in the order of 100 nanoliters in volume. This means that 10-fold less protein is used per-experiment when compared to crystallization trials setup by hand (in the order of 1 microliter).

Several factors are known to inhibit or mar crystallization. The growing crystals are generally held at a constant temperature and protected from shocks or vibrations that might disturb their crystallization. Impurities in the molecules or in the crystallization solutions are often inimical to crystallization. Conformational flexibility in the molecule also tends to make crystallization less likely, due to entropy. Ironically, molecules that tend to self-assemble into regular helices are often unwilling to assemble into crystals. Crystals can be marred by twinning, which can occur when a unit cell can pack equally favorably in multiple orientations; although recent advances in computational methods may allow solving the structure of some twinned crystals. Having failed to crystallize a target molecule, a crystallographer may try again with a slightly modified version of the molecule; even small changes in molecular properties can lead to large differences in crystallization behavior.

Data collection

Mounting the crystal

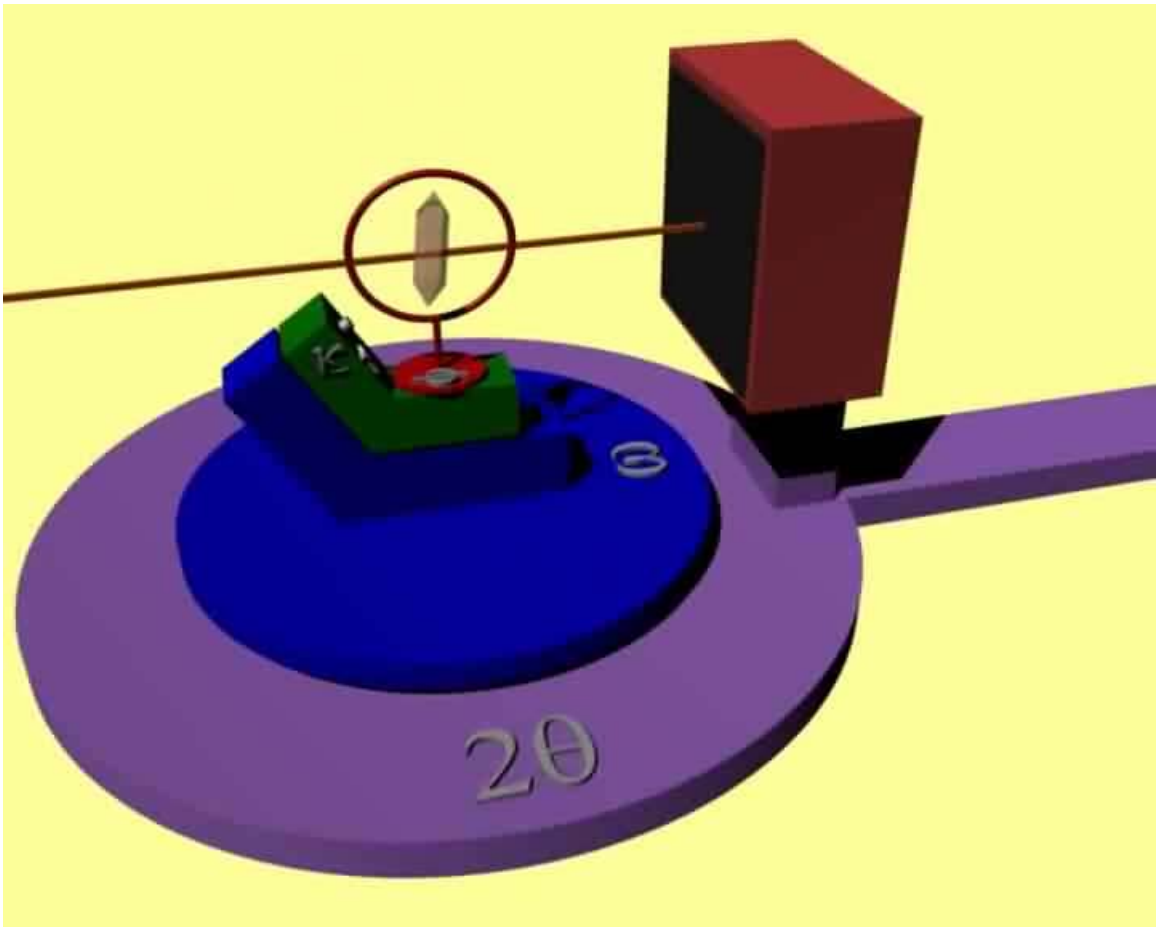


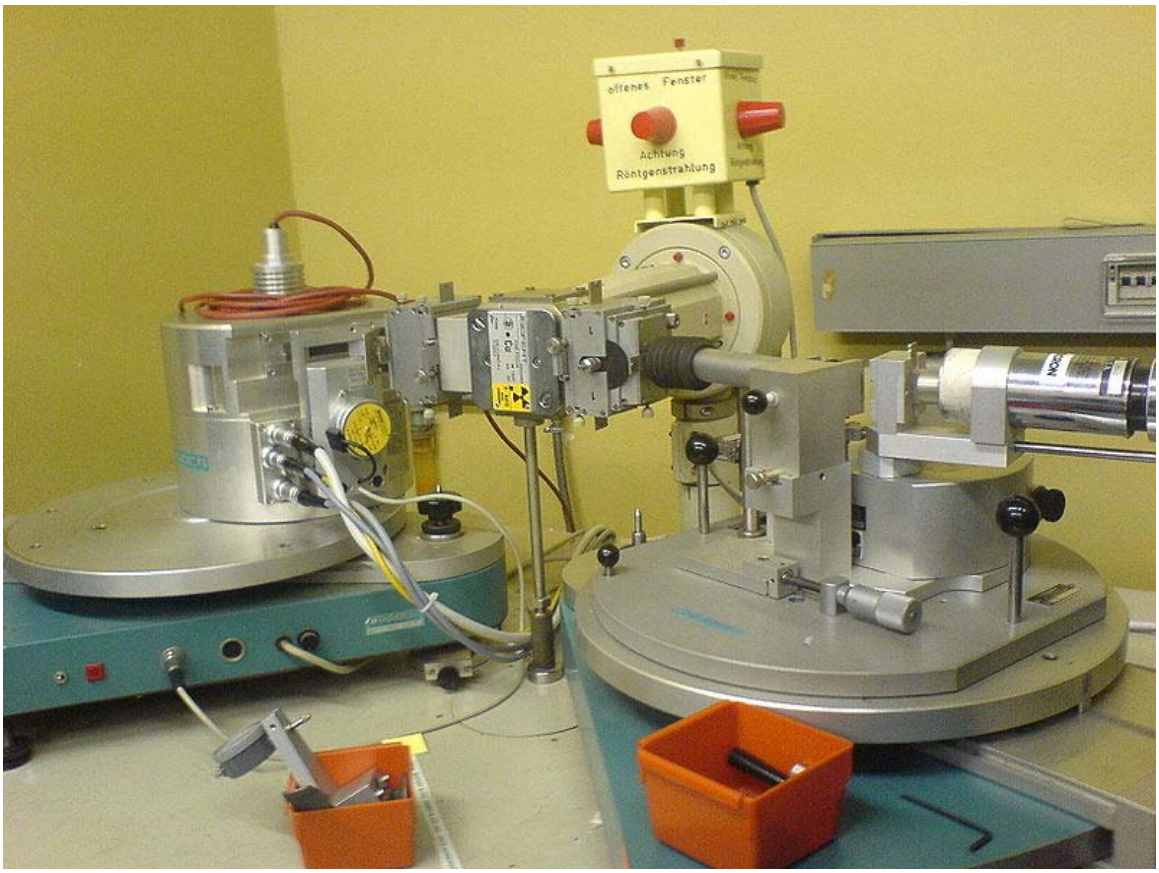
Image showing the five motions possible with a four-circle kappa goniometer. The rotations about each of the four angles φ, κ, ω and 2θ leave the crystal within the X-ray beam, but change the crystal orientation. The detector (red box) can be slid closer or further away from the crystal, allowing higher resolution data to be taken (if closer) or better discernment of the Bragg peaks (if further away).

The crystal is mounted for measurements so that it may be held in the X-ray beam and rotated. There are several methods of mounting. Although crystals were once loaded into glass capillaries with the crystallization solution (the mother liquor), a modern approach is to scoop the crystal up in a tiny loop, made of nylon or plastic and attached to a solid rod, that is then flash-frozen with liquid nitrogen. This freezing reduces the radiation damage of the X-rays, as well as the noise in the Bragg peaks due to thermal motion (the Debye-Waller effect). However, untreated crystals often crack if flash-frozen; therefore, they are generally pre-soaked in a cryoprotectant solution before freezing. Unfortunately, this pre-soak may itself cause the crystal to crack, ruining it for crystallography. Generally, successful cryo-conditions are identified by trial and error.

The capillary or loop is mounted on a goniometer, which allows it to be positioned accurately within the X-ray beam and rotated. Since both the crystal and the beam are often very small, the crystal must be centered within the beam to within ~25 micrometers accuracy, which is aided by a camera focused on the crystal. The most common type of goniometer is the "kappa goniometer", which offers three angles of rotation: the ω angle, which rotates about an axis perpendicular to the beam; the κ angle, about an axis at ~50° to the ω axis; and, finally, the φ angle about the loop/capillary axis. When the κ angle is zero, the ω and φ axes are aligned. The κ rotation allows for convenient mounting of the crystal, since the arm in which the crystal is mounted may be swung out towards the crystallographer. The oscillations carried out during data collection (mentioned below) involve the ω axis only. An older type of goniometer is the four-circle goniometer, and its relatives such as the six-circle goniometer.

X-ray sources

The mounted crystal is then irradiated with a beam of monochromatic X-rays. The brightest and most useful X-ray sources are synchrotrons; their much higher luminosity allows for better resolution. They also make it convenient to tune the wavelength of the radiation, which is useful for multi-wavelength anomalous dispersion (MAD) phasing, described below. Synchrotrons are generally national facilities, each with several dedicated beamlines where data is collected around the clock, seven days a week.

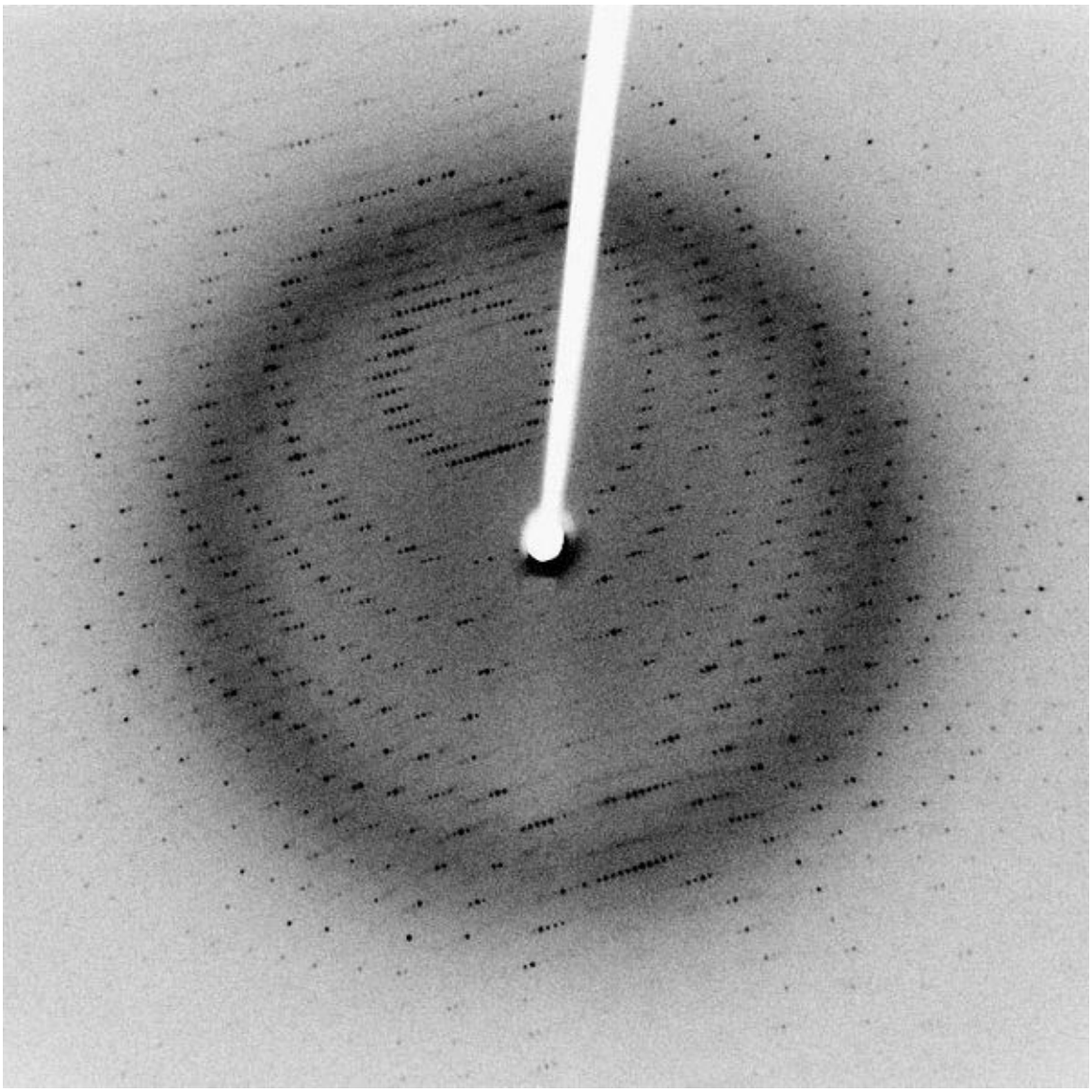


A diffractometer

Smaller, X-ray generators are often used in laboratories to check the quality of crystals before bringing them to a synchrotron and sometimes to solve a crystal structure. In such systems, electrons are boiled off of a cathode and accelerated through a strong electric potential of ~50 kV; having reached a high speed, the electrons collide with a metal plate, emitting *bremsstrahlung* and some strong spectral lines corresponding to the excitation of inner-shell electrons of the metal. The most common metal used is copper, which can be kept cool easily, due to its high thermal conductivity, and which produces strong K_α and K_β lines. The K_β line is sometimes suppressed with a thin (~10 μm) nickel foil. The simplest and cheapest variety of sealed X-ray tube has a stationary anode (the Crookes tube) and produces ~2 kW of X-ray radiation. The more expensive variety has a rotating-anode type source that produces ~14 kW of X-ray radiation.

X-rays are generally filtered (by use of X-Ray Filters) to a single wavelength (made monochromatic) and collimated to a single direction before they are allowed to strike the crystal. The filtering not only simplifies the data analysis, but also removes radiation that degrades the crystal without contributing useful information. Collimation is done either with a collimator (basically, a long tube) or with a clever arrangement of gently curved mirrors. Mirror systems are preferred for small crystals (under 0.3 mm) or with large unit cells (over 150 Å)

Recording the reflections



An X-ray diffraction pattern of a crystallized enzyme. The pattern of spots (called *reflections*) can be used to determine the structure of the enzyme.

When a crystal is mounted and exposed to an intense beam of X-rays, it scatters the X-rays into a pattern of spots or *reflections* that can be observed on a screen behind the crystal. A similar pattern may be seen by shining a laser pointer at a compact disc. The relative intensities of these spots provide the information to determine the arrangement of molecules within the crystal in atomic detail. The intensities of these reflections may be recorded with photographic film, an area detector or with a charge-coupled device (CCD) image sensor. The peaks at small angles correspond to low-resolution data, whereas those at high angles represent high-resolution data; thus, an upper limit on the eventual resolution of the structure can be determined from the first few images. Some measures

of diffraction quality can be determined at this point, such as the mosaicity of the crystal and its overall disorder, as observed in the peak widths. Some pathologies of the crystal that would render it unfit for solving the structure can also be diagnosed quickly at this point.

One image of spots is insufficient to reconstruct the whole crystal; it represents only a small slice of the full Fourier transform. To collect all the necessary information, the crystal must be rotated step-by-step through 180°, with an image recorded at every step; actually, slightly more than 180° is required to cover reciprocal space, due to the curvature of the Ewald sphere. However, if the crystal has a higher symmetry, a smaller angular range such as 90° or 45° may be recorded. The rotation axis should be changed at least once, to avoid developing a "blind spot" in reciprocal space close to the rotation axis. It is customary to rock the crystal slightly (by 0.5-2°) to catch a broader region of reciprocal space.

Multiple data sets may be necessary for certain phasing methods. For example, MAD phasing requires that the scattering be recorded at least three (and usually four, for redundancy) wavelengths of the incoming X-ray radiation. A single crystal may degrade too much during the collection of one data set, owing to radiation damage; in such cases, data sets on multiple crystals must be taken.

Data analysis

Crystal symmetry, unit cell, and image scaling

The recorded series of two-dimensional diffraction patterns, each corresponding to a different crystal orientation, is converted into a three-dimensional model of the electron density; the conversion uses the mathematical technique of Fourier transforms, which is explained below. Each spot corresponds to a different type of variation in the electron density; the crystallographer must determine *which* variation corresponds to *which* spot (*indexing*), the relative strengths of the spots in different images (*merging and scaling*) and how the variations should be combined to yield the total electron density (*phasing*).

Data processing begins with *indexing* the reflections. This means identifying the dimensions of the unit cell and which image peak corresponds to which position in reciprocal space. A byproduct of indexing is to determine the symmetry of the crystal, i.e., its *space group*. Some space groups can be eliminated from the beginning. For example, reflection symmetries cannot be observed in chiral molecules; thus, only 65 space groups of 243 possible are allowed for protein molecules which are almost always chiral. Indexing is generally accomplished using an *autoindexing* routine. Having assigned symmetry, the data is then *integrated*. This converts the hundreds of images containing the thousands of reflections into a single file, consisting of (at the very least) records of the Miller index of each reflection, and an intensity for each reflection (at this state the file often also includes error estimates and measures of partiality (what part of a given reflection was recorded on that image)).

A full data set may consist of hundreds of separate images taken at different orientations of the crystal. The first step is to merge and scale these various images, that is, to identify which peaks appear in two or more images (*merging*) and to scale the relative images so that they have a consistent intensity scale. Optimizing the intensity scale is critical because the relative intensity of the peaks is the key information from which the structure is determined. The repetitive technique of crystallographic data collection and the often high symmetry of crystalline materials cause the diffractometer to record many symmetry-equivalent reflections multiple times. This allows calculating the symmetry-related R-factor, a reliability index based upon how similar are the measured intensities of symmetry-equivalent reflections, thus assessing the quality of the data.

Initial phasing

The data collected from a diffraction experiment is a reciprocal space representation of the crystal lattice. The position of each diffraction 'spot' is governed by the size and shape of the unit cell, and the inherent symmetry within the crystal. The intensity of each diffraction 'spot' is recorded, and this intensity is proportional to the square of the *structure factor* amplitude. The structure factor is a complex number containing information relating to both the amplitude and phase of a wave. In order to obtain an interpretable *electron density map*, both amplitude and phase must be known (an electron density map allows a crystallographer to build a starting model of the molecule). The phase cannot be directly recorded during a diffraction experiment: this is known as the phase problem. Initial phase estimates can be obtained in a variety of ways:

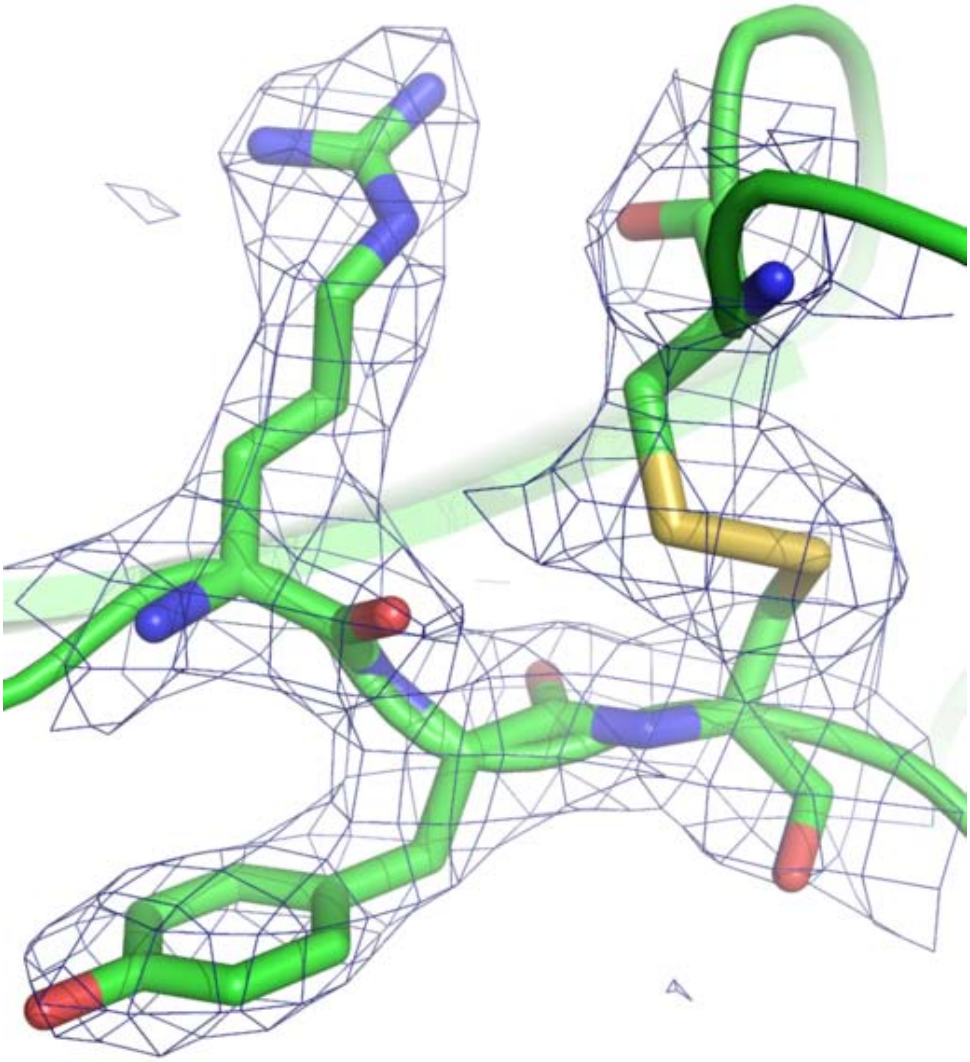
- ***Ab initio*** **phasing** or **direct methods** - This is usually the method of choice for small molecules (<1000 non-hydrogen atoms), and has been used successfully to solve the phase problems for small proteins. If the resolution of the data is better than 1.4 Å (140 pm), direct methods can be used to obtain phase information, by exploiting known phase relationships between certain groups of reflections.

- **Molecular replacement** - if a related structure is known, it can be used as a search model in molecular replacement to determine the orientation and position of the molecules within the unit cell. The phases obtained this way can be used to generate *electron density maps*.

- **Anomalous X-ray scattering** (*MAD or SAD phasing*) - the X-ray wavelength may be scanned past an absorption edge of an atom, which changes the scattering in a known way. By recording full sets of reflections at three different wavelengths (far below, far above and in the middle of the absorption edge) one can solve for the substructure of the anomalously diffracting atoms and thence the structure of the whole molecule. The most popular method of incorporating anomalous scattering atoms into proteins is to express the protein in a methionine auxotroph (a host incapable of synthesizing methionine) in a media rich in seleno-methionine, which contains selenium atoms. A MAD experiment can then be conducted around the absorption edge, which should then yield the position of any methionine residues within the protein, providing initial phases.

- **Heavy atom methods** (multiple isomorphous replacement) - If electron-dense metal atoms can be introduced into the crystal, direct methods or Patterson-space methods can be used to determine their location and to obtain initial phases. Such heavy atoms can be introduced either by soaking the crystal in a heavy atom-containing solution, or by co-crystallization (growing the crystals in the presence of a heavy atom). As in MAD phasing, the changes in the scattering amplitudes can be interpreted to yield the phases. Although this is the original method by which protein crystal structures were solved, it has largely been superseded by MAD phasing with selenomethionine.

Model building and phase refinement



A protein crystal structure at 2.7 Å resolution. The mesh encloses the region in which the electron density exceeds a given threshold. The straight segments represent chemical bonds between the non-hydrogen atoms of an arginine (upper left), a tyrosine (lower left),

a disulfide bond (upper right, in yellow), and some peptide groups (running left-right in the middle). The two curved green tubes represent spline fits to the polypeptide backbone.

Having obtained initial phases, an initial model can be built. This model can be used to refine the phases, leading to an improved model, and so on. Given a model of some atomic positions, these positions and their respective Debye-Waller factors (or **B**-factors, accounting for the thermal motion of the atom) can be refined to fit the observed diffraction data, ideally yielding a better set of phases. A new model can then be fit to the new electron density map and a further round of refinement is carried out. This continues until the correlation between the diffraction data and the model is maximized. The agreement is measured by an *R*-factor defined as

$$R = \frac{\sum_{\text{all reflections}} |F_o - F_c|}{\sum_{\text{all reflections}} |F_o|}$$

where *F* is the structure factor. A similar quality criterion is R_{free}, which is calculated from a subset (~10%) of reflections that were not included in the structure refinement. Both *R* factors depend on the resolution of the data. As a rule of thumb, R_{free} should be approximately the resolution in Ångströms divided by 10; thus, a data-set with 2 Å resolution should yield a final $R_{\text{free}} \sim 0.2$. Chemical bonding features such as stereochemistry, hydrogen bonding and distribution of bond lengths and angles are complementary measures of the model quality. Phase bias is a serious problem in such iterative model building. *Omit maps* are a common technique used to check for this.

It may not be possible to observe every atom of the crystallized molecule - it must be remembered that the resulting electron density is an average of all the molecules within the crystal. In some cases, there is too much residual disorder in those atoms, and the resulting electron density for atoms existing in many conformations is smeared to such an extent that it is no longer detectable in the electron density map. Weakly scattering atoms such as hydrogen are routinely invisible. It is also possible for a single atom to appear multiple times in an electron density map, e.g., if a protein sidechain has multiple (<4) allowed conformations. In still other cases, the crystallographer may detect that the covalent structure deduced for the molecule was incorrect, or changed. For example, proteins may be cleaved or undergo post-translational modifications that were not detected prior to the crystallization.

Deposition of the structure

Once the model of a molecule's structure has been finalized, it is often deposited in a crystallographic database such as the Cambridge Structural Database (for small molecules), the Inorganic Crystal Structure Database (ICSD) (for inorganic compounds) or the Protein Data Bank (for protein structures). Many structures obtained in private commercial ventures to crystallize medicinally relevant proteins, are not deposited in public crystallographic databases.

Diffraction theory

The main goal of X-ray crystallography is to determine the density of electrons $f(\mathbf{r})$ throughout the crystal, where $\mathbf{r}$ represents the three-dimensional position vector within the crystal. To do this, X-ray scattering is used to collect data about its Fourier transform $F(\mathbf{q})$, which is inverted mathematically to obtain the density defined in real space, using the formula

$$f(\mathbf{r}) = \int \frac{d\mathbf{q}}{(2\pi)^3} F(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}}$$

where the integral is taken over all values of $\mathbf{q}$. The three-dimensional real vector $\mathbf{q}$ represents a point in reciprocal space, that is, to a particular oscillation in the electron density as one moves in the direction in which $\mathbf{q}$ points. The length of $\mathbf{q}$ corresponds to 2π divided by the wavelength of the oscillation. The corresponding formula for a Fourier transform will be used below

$$F(\mathbf{q}) = \int d\mathbf{r} f(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}}$$

where the integral is summed over all possible values of the position vector $\mathbf{r}$ within the crystal.

The Fourier transform $F(\mathbf{q})$ is generally a complex number, and therefore has a magnitude $|F(\mathbf{q})|$ and a phase $\varphi(\mathbf{q})$ related by the equation

$$F(\mathbf{q}) = |F(\mathbf{q})| e^{i\phi(\mathbf{q})}$$

The intensities of the reflections observed in X-ray diffraction give us the magnitudes $|F(\mathbf{q})|$ but not the phases $\varphi(\mathbf{q})$. To obtain the phases, full sets of reflections are collected with known alterations to the scattering, either by modulating the wavelength past a certain absorption edge or by adding strongly scattering (i.e., electron-dense) metal atoms such as mercury. Combining the magnitudes and phases yields the full Fourier transform $F(\mathbf{q})$, which may be inverted to obtain the electron density $f(\mathbf{r})$.

Crystals are often idealized as being *perfectly* periodic. In that ideal case, the atoms are positioned on a perfect lattice, the electron density is perfectly periodic, and the Fourier transform $F(\mathbf{q})$ is zero except when $\mathbf{q}$ belongs to the reciprocal lattice (the so-called *Bragg peaks*). In reality, however, crystals are not perfectly periodic; atoms vibrate about their mean position, and there may be disorder of various types, such as mosaicity, dislocations, various point defects, and heterogeneity in the conformation of crystallized molecules. Therefore, the Bragg peaks have a finite width and there may be significant *diffuse scattering*, a continuum of scattered X-rays that fall between the Bragg peaks.

Intuitive understanding by Bragg's law

An intuitive understanding of X-ray diffraction can be obtained from the Bragg model of diffraction. In this model, a given reflection is associated with a set of evenly spaced sheets running through the crystal, usually passing through the centers of the atoms of the crystal lattice. The orientation of a particular set of sheets is identified by its three Miller indices (*h*, *k*, *l*), and let their spacing be noted by *d*. William Lawrence Bragg proposed a model in which the incoming X-rays are scattered specularly (mirror-like) from each plane; from that assumption, X-rays scattered from adjacent planes will combine constructively (constructive interference) when the angle θ between the plane and the X-ray results in a path-length difference that is an integer multiple *n* of the X-ray wavelength λ.

$$2d \sin\theta = n\lambda$$

A reflection is said to be *indexed* when its Miller indices (or, more correctly, its reciprocal lattice vector components) have been identified from the known wavelength and the scattering angle 2θ. Such indexing gives the unit-cell parameters, the lengths and angles of the unit-cell, as well as its space group. Since Bragg's law does not interpret the relative intensities of the reflections, however, it is generally inadequate to solve for the arrangement of atoms within the unit-cell; for that, a Fourier transform method must be carried out.

Scattering as a Fourier transform

The incoming X-ray beam has a polarization and should be represented as a vector wave; however, for simplicity, let it be represented here as a scalar wave. We also ignore the complication of the time dependence of the wave and just focus on the wave's spatial dependence. Plane waves can be represented by a wave vector $\mathbf{k}_{in}$, and so the strength of the incoming wave at time *t=0* is given by

$$Ae^{i\mathbf{k}_{in}\cdot\mathbf{r}}$$

At position **r** within the sample, let there be a density of scatterers *f*(**r**); these scatterers should produce a scattered spherical wave of amplitude proportional to the local amplitude of the incoming wave times the number of scatterers in a small volume *dV* about **r**

$$\text{amplitude of scattered wave} = Ae^{i\mathbf{k}\cdot\mathbf{r}}Sf(\mathbf{r})dV$$

where *S* is the proportionality constant.

Let's consider the fraction of scattered waves that leave with an outgoing wave-vector of $\mathbf{k}_{out}$ and strike the screen at $\mathbf{r}_{screen}$. Since no energy is lost (elastic, not inelastic scattering), the wavelengths are the same as are the magnitudes of the wave-vectors $|\mathbf{k}_{in}|=|\mathbf{k}_{out}|$. From

the time that the photon is scattered at **r** until it is absorbed at $\mathbf{r}_{\text{screen}}$, the photon undergoes a change in phase

$$e^{i\mathbf{k}_{out}\cdot(\mathbf{r}_{\text{screen}}-\mathbf{r})}$$

The net radiation arriving at $\mathbf{r}_{\text{screen}}$ is the sum of all the scattered waves throughout the crystal

$$AS\int d\mathbf{r} f(\mathbf{r}) e^{i\mathbf{k}_{in}\cdot\mathbf{r}} e^{i\mathbf{k}_{out}\cdot(\mathbf{r}_{\text{screen}}-\mathbf{r})} = ASe^{i\mathbf{k}_{out}\cdot\mathbf{r}_{\text{screen}}}\int d\mathbf{r} f(\mathbf{r}) e^{i(\mathbf{k}_{in}-\mathbf{k}_{out})\cdot\mathbf{r}}$$

which may be written as a Fourier transform

$$ASe^{i\mathbf{k}_{out}\cdot\mathbf{r}_{\text{screen}}}\int d\mathbf{r} f(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} = ASe^{i\mathbf{k}_{out}\cdot\mathbf{r}_{\text{screen}}} F(\mathbf{q})$$

where $\mathbf{q} = \mathbf{k}_{\text{out}} - \mathbf{k}_{\text{in}}$. The measured intensity of the reflection will be square of this amplitude

$$A^2S^2\left|F(\mathbf{q})\right|^2$$

Friedel and Bijvoet mates

For every reflection corresponding to a point **q** in the reciprocal space, there is another reflection of the same intensity at the opposite point **-q**. This opposite reflection is known as the *Friedel mate* of the original reflection. This symmetry results from the mathematical fact that the density of electrons $f(\mathbf{r})$ at a position **r** is always a real number. As noted above, $f(\mathbf{r})$ is the inverse transform of its Fourier transform $F(\mathbf{q})$; however, such an inverse transform is a complex number in general. To ensure that $f(\mathbf{r})$ is real, the Fourier transform $F(\mathbf{q})$ must be such that the Friedel mates $F(-\mathbf{q})$ and $F(\mathbf{q})$ are complex conjugates of one another. Thus, $F(-\mathbf{q})$ has the same magnitude as $F(\mathbf{q})$ but they have the opposite phase, i.e., $\varphi(\mathbf{q}) = -\varphi(\mathbf{q})$

$$F(-\mathbf{q}) = \left|F(-\mathbf{q})\right| e^{i\phi(-\mathbf{q})} = F^*(\mathbf{q}) = \left|F(\mathbf{q})\right| e^{-i\phi(\mathbf{q})}$$

The equality of their magnitudes ensures that the Friedel mates have the same intensity $|F|^2$. This symmetry allows one to measure the full Fourier transform from only half the reciprocal space, e.g., by rotating the crystal slightly more than a 180°, instead of a full turn. In crystals with significant symmetry, even more reflections may have the same intensity (Bijvoet mates); in such cases, even less of the reciprocal space may need to be measured, e.g., slightly more than 90°.

The Friedel-mate constraint can be derived from the definition of the inverse Fourier transform

$$f(\mathbf{r}) = \int \frac{d\mathbf{q}}{(2\pi)^3} F(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}} = \int \frac{d\mathbf{q}}{(2\pi)^3} |F(\mathbf{q})| e^{i\phi(\mathbf{q})} e^{i\mathbf{q}\cdot\mathbf{r}}$$

Since Euler's formula states that $e^{ix} = \cos(x) + i\sin(x)$, the inverse Fourier transform can be separated into a sum of a purely real part and a purely imaginary part

$$f(\mathrm{r}) = \int \frac{d\mathrm{q}}{(2\pi)^3} |F(\mathrm{q})| e^{i(\phi+\mathrm{q}\cdot\mathrm{r})} = \int \frac{d\mathrm{q}}{(2\pi)^3} |F(\mathrm{q})| \cos(\phi+\mathrm{q}\cdot\mathrm{r}) + i\int \frac{d\mathrm{q}}{(2\pi)^3} |F(\mathrm{q})| \sin(\phi+\mathrm{q}\cdot\mathrm{r}) = I_{\cos} + iI_{\sin}$$

The function $f(\mathbf{r})$ is real if and only if the second integral $I_{\sin}$ is zero for all values of $\mathbf{r}$. In turn, this is true if and only if the above constraint is satisfied

$$I_{\sin} = \int \frac{d\mathbf{q}}{(2\pi)^3} |F(\mathbf{q})| \sin(\phi + \mathbf{q}\cdot\mathbf{r}) = \int \frac{d\mathbf{q}}{(2\pi)^3} |F(-\mathbf{q})| \sin(-\phi - \mathbf{q}\cdot\mathbf{r}) = -I_{\sin}$$

since $I_{\sin} = -I_{\sin}$ implies that $I_{\sin}=0$.

Ewald's sphere

Each X-ray diffraction image represents only a slice, a spherical slice of reciprocal space, as may be seen by the Ewald sphere construction. Both $\mathbf{k}_{\text{out}}$ and $\mathbf{k}_{\text{in}}$ have the same length, due to the elastic scattering, since the wavelength has not changed. Therefore, they may be represented as two radial vectors in a sphere in reciprocal space, which shows the values of $\mathbf{q}$ that are sampled in a given diffraction image. Since there is a slight spread in the incoming wavelengths of the incoming X-ray beam, the values of$|F(\mathbf{q})|$can be measured only for $\mathbf{q}$ vectors located between the two spheres corresponding to those radii. Therefore, to obtain a full set of Fourier transform data, it is necessary to rotate the crystal through slightly more than 180°, or sometimes less if sufficient symmetry is present. A full 360° rotation is not needed because of a symmetry intrinsic to the Fourier transforms of real functions (such as the electron density), but "slightly more" than 180° is needed to cover all of reciprocal space within a given resolution because of the curvature of the Ewald sphere. In practice, the crystal is rocked by a small amount (0.25-1°) to incorporate reflections near the boundaries of the spherical Ewald shells.

Patterson function

A well-known result of Fourier transforms is the autocorrelation theorem, which states that the autocorrelation $c(\mathbf{r})$ of a function $f(\mathbf{r})$

$$c(\mathbf{r}) = \int d\mathbf{x} f(\mathbf{x}) f(\mathbf{x}+\mathbf{r}) = \int \frac{d\mathbf{q}}{(2\pi)^3} C(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}}$$

has a Fourier transform $C(\mathbf{q})$ that is the squared magnitude of $F(\mathbf{q})$

$$C(\mathbf{q}) = |F(\mathbf{q})|^2$$

Therefore, the autocorrelation function $c(\mathbf{r})$ of the electron density (also known as the *Patterson function*) can be computed directly from the reflection intensities, without computing the phases. In principle, this could be used to determine the crystal structure directly; however, it is difficult to realize in practice. The autocorrelation function corresponds to the distribution of vectors between atoms in the crystal; thus, a crystal of N atoms in its unit cell may have $N(N-1)$ peaks in its Patterson function. Given the inevitable errors in measuring the intensities, and the mathematical difficulties of reconstructing atomic positions from the interatomic vectors, this technique is rarely used to solve structures, except for the simplest crystals.

Advantages of a crystal

In principle, an atomic structure could be determined from applying X-ray scattering to non-crystalline samples, even to a single molecule. However, crystals offer a much stronger signal due to their periodicity. A crystalline sample is by definition periodic; a crystal is composed of many unit cells repeated indefinitely in three independent directions. Such periodic systems have a Fourier transform that is concentrated at periodically repeating points in reciprocal space known as *Bragg peaks*; the Bragg peaks correspond to the reflection spots observed in the diffraction image. Since the amplitude at these reflections grows linearly with the number N of scatterers, the observed *intensity* of these spots should grow quadratically, like N^2. In other words, using a crystal concentrates the weak scattering of the individual unit cells into a much more powerful, coherent reflection that can be observed above the noise. This is an example of constructive interference.

In a liquid, powder or amorphous sample, molecules within that sample are in random orientations. Such samples have a continuous Fourier spectrum that uniformly spreads its amplitude thereby reducing the measured signal intensity, as is observed in SAXS. More importantly, the orientational information is lost. Although theoretically possible, it is experimentally difficult to obtain atomic-resolution structures of complicated, asymmetric molecules from such rotationally averaged data. An intermediate case is fiber diffraction in which the subunits are arranged periodically in at least one dimension.