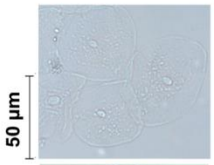
# BIOLOGISCHE ANALYTIK

## 1. PRINCIPLES OF MICROSCOPY

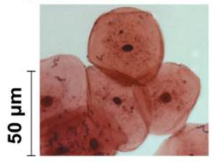
### OVERVIEW

#### Light microscopy

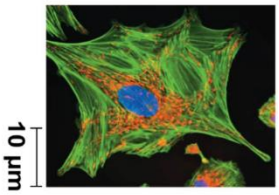
**Brightfield (unstained specimen).** Light passes directly through the specimen (human cheek epithelial cells). Unless the cell is naturally pigmented or artificially stained, the image has little contrast.



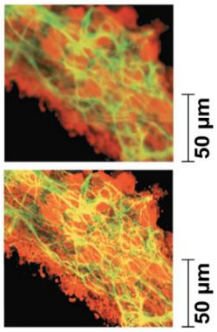

**Brightfield (stained specimen).** Staining with various dyes enhances contrast. Most staining procedures require that cells be fixed (preserved), thereby killing them.



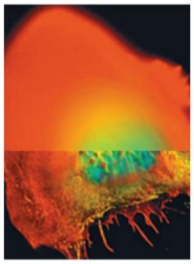

**Fluorescence.** The locations of specific molecules in the cell can be revealed by labeling the molecules with fluorescent dyes or antibodies; some cells have molecules that fluoresce on their own. Fluorescent substances absorb ultraviolet radiation and emit visible light. In this fluorescently labeled uterine cell, the nucleus is blue, mitochondria are orange and the cytoskeleton is green.



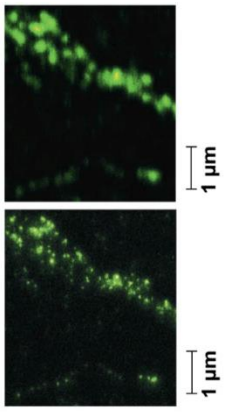

**Confocal.** The top image is a standard fluorescence micrograph of of fluorescently labeled nervous tissue (neurons are green, support cells are orange, and regions of overlap are yellow); below is a confocal image of the same tissue. Using a laser, this 'optical sectioning' technique eliminates out-of-focus light from a thick sample, creating a single plane of fluorescence in the image. By capturing sharp images at many different planes, a 3D reconstruction can be created. The standard image is blurry, because out-of-focus light is not excluded.




**Deconvolution.** The top of this split image is a compilation of standard fluorescence micrographs through the depth of a white blood cell. Below is an image of the same cell reconstructed from many blurry images at different planes, each of which was processed using deconvolution software. This process digitally removes out-of-focus light and reassigns it to its source, creating a much sharper 3D image.



**Super-resolution.** On the top is a confocal image of part of a neuron, using a fluorescent label that binds to a molecule clustered in vesicles that are 40nm in diameter. The greenish-yellow spots are blurry, because 40nm is below the 200-nm limit of resolution for standard light microscopy. Below is an image of the same part of the cell, seen using a new super-resolution technique. Sophisticated equipment is used to light up individual fluorescent molecules and record their position. Combining information from many molecules in different planes "breaks" the limit of resolution, resulting in the sharp greenish-yellow dots seen here. (Each dot is a 40-nm vesicle.)

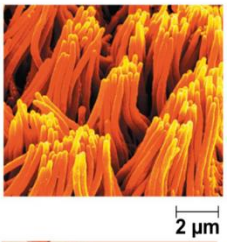



### FLUORESCENCE MICROSCOPY

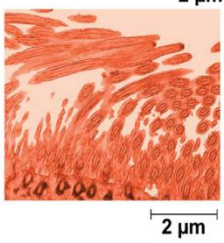
- wide-field (i.e. epifluorescence)
  - inherently faster than a confocal microscope →entire view area of the sample is illuminated at the same time and light emitted from the entire area is recorded simultaneously on an area
- deconvolution microscopy
  - uses images recorded with a conventional wide-field fluorescent microscope and enhances these images computationally → mathematical procedure of deconvolution is used to identify & reduce blurring induced by light from out-of-focus planes of the sample
- confocal microscopy
  - builds up the image by scanning a very small measurement spot across the sample
  - advantage of confocal microscopy relative to wide-field fluorescence microscopy is particularly pronounced in thick samples
- light sheet microscopy
  - light sensitive sample; illuminating only that slice of the sample, from which data is currently being collected (vs. epifluorescence illumination where light traverses and causes deleterious effects in the entire sample volume)

#### Electron microscopy

**Scanning electron microscopy (SEM).** Micrographs taken with a SEM show a 3D image of the surface of a specimen. This SEM shows the surface of a cell from a trachea covered with cilia. Electron micrographs are black and white, but are often artificially colorized to highlight particular structures.

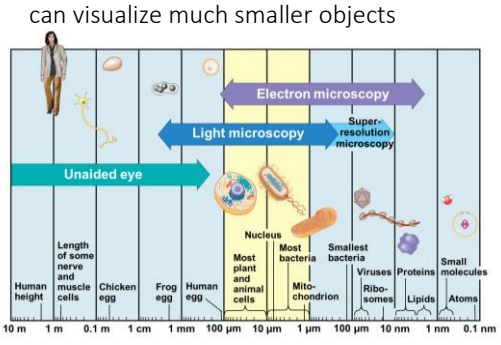



**Transmission electron microscopy (TEM).** A TEM profiles a thin section of a specimen. Here we see a section through a tracheal cell, revealing its internal structure. In preparing the specimen, some cilia were cut along their lengths, creating longitudinal sections, while other cilia were cut straight across, creating cross section.




### TWO TYPES OF MICROSCOPY

- Light microscopy
  - Objects are examined by **shining light through a sample** and measuring its Interaction with objects
  - obtain information about the specimen from the **transmission and diffraction pattern of the light**
  - Limitation: only visualize relatively large structures
  - **fluorescence microscopy** (good to focus only on certain structures of a cell by highlighting them)
- electron microscopy
  - objects are **probed by electrons**
  - can visualize much smaller objects



# PRINCIPLES OF LIGHT MICROSCOPY

## BASIC PROPERTIES OF LIGHT

Light is a form of energy called electromagnetic radiation. This energy is contained in discrete units or quanta called photons that have the properties of both particles and waves

- Light is **energy** in form of an **electromagnetic wave**
- amount of **energy light** contains is represented **by its wavelength** (not by its speed → light has no mass → travels always at the speed of light)
- The smaller the wavelength, the higher its energy
- Light is the smallest form of energy that can be transported (photon)
- Has partly **wave-character**, partly **matter-character** (**wave-particle-duality**)
- can be **reflected, absorbed or transmitted**
- **is created when electrons fall** from a high-energy state into a lower one → energy won is emitted in form of **radiation**
- Moving such charged electrons causes **electric fields** that again induce **magnetic fields**
- **Visible** (white) light = small range of the electromagnetic spectrum (400-700nm)
  - Only visible light **propagates well in water**

### IN MICROSCOPY WE USE:

- White light = contains a **combination of wavelengths**, often the entire spectrum of visible light (400-700nm)
  - Monochromatic light = visible light of **only one wavelength** (laser as light source → very important for fluorescence microscopy → selective agitation)

### DIFFERENT FORMS OF LIGHT:

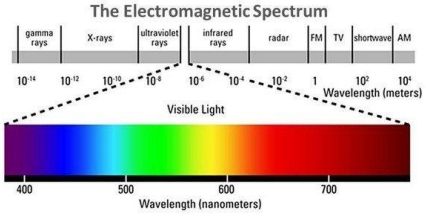
there are **different representations** to describe different aspects of light behavior (but are the same thing!)

- **light as a wave** (describing how light travels through space, interference)
  → Light has wave-like properties: **Diffraction**
- **light as rays** (how light interacts objects))
- **light as photons** → elementary particle of light
  → quanta of electromagnetic radiation (excitation of electrons through absorption of light)

- Light possesses both particle and wave properties, **regardless of the medium through which it travels**

### WAVELENGHT:

- All the different **wavelength are the same but differ in their frequency**




- **Gamma Rays**: smallest wavelength, because they're the highest energy Photons
- The longer the wavelength of light, the lower its frequency and the lower the energy of its photons

### PHOTONS:

- the wavelength of the photon gives the energy that is contained in the photon
- the shorter the wavelenght the higher the energy

$c = \lambda * v$ (= 0.3 *$10^6$ m/s)

$$E(Photon) = h * v = h * \frac{c}{\lambda}$$

$h$ = Planck constant = (6.62607004 *$10^{-34}$ $m^2$ kg/s)
$v$ = frequency of oscillation
$\lambda$ = Wavelenght [nm]
Avrogado's Number: 6.022 *$10^{23}$ → needed to convert the energy of a mol of photons into the energy of a single photon

What is the wavelength (in nm) of light, when one mole of its photons carries 160 kJ (= 38 kcal)

$$E_{Photon} = \frac{160'000J}{6.022 * 10^{23}} = 2.66 * 10^{-19} J$$

$$E = \frac{hc}{\lambda} \rightarrow \lambda = \frac{hc}{E} = \frac{6.62*10^{-34} Js * 3*10^8 \frac{m}{s}}{2.66*10^{-19} J} = 7.5 * 10^{-7} \text{m}$$

= 750nm

Wavelength: Blue: 63 kcal/mol, red 42kcal/mol
→ C-C-Bindung trennen: 100kcal/mol

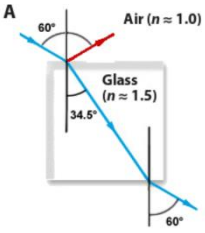
The energy carried by photons of visible light approach the energies required for rupturing chemical bonds. In light microscopy we must therefore be aware of the possibility that the light used for viewing may cause artefacts in the sample

An increase in light intensity changes the number of photon that hit the sample, but the energy of each photon stays the same

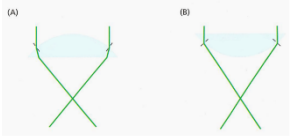
## REFLECTION AND REFRACTION OF LIGHT

The **fastest** "matter" known, since it has no matter
**c = light speed = 2.8x$10^6$ m/s → only in Vacuum!**

- Light does **not change its properties** when entering another medium → but **it changes speed**
- In materials the speed of light is lower!
- refractive index n = factor by which light travels slower in a material (= Brechungsindex)
- absorbance: As light passes through any kind of matter (including optical lenses) a part of it will be absorbed
- reflection: reflected rays leave the surface in the same angle as they hit it
- refraction (if medium changes): rays do not leave the same direction/angle as they entered it
  → Light passing from one type of matter to another is **refracted only when refractive index of the two is different**
- When light enters a medium with a different optical density, it bends in a direction that depends on the refractive indices of the two media



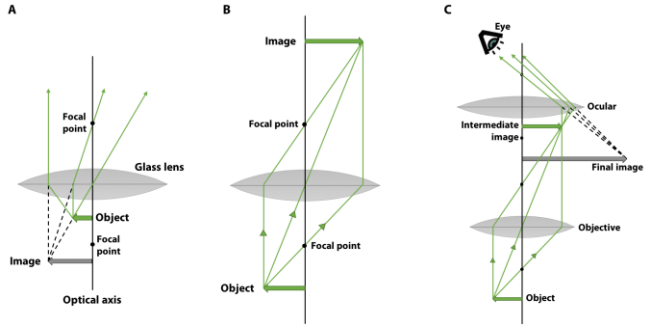

- when traveling from a **less dense (lower refractive index)** to a **denser** medium, the rays/light ist refracted **towards the normal line** (90° angle to surface)
- Air (n=1), glass (n=1.5 → denser)




light rays will converge (be focused)

## LENSES – MICROSCOPY

- compound microscope = device with multiple lenses (reaches higher magnifications)
- **two convex lenses** aligned in series:
  - object glass close to the specimen
  - eyepiece (ocular) lens close to the observer's eye
- When the sample is located outside of the focal length of the lens, the lens creates a magnified image behind itself




A) distance object-lense < distance focal lenght-lense
→ magnifying glass (Verkleinerung)
→ virtual image as light rays do not actually pass through the object

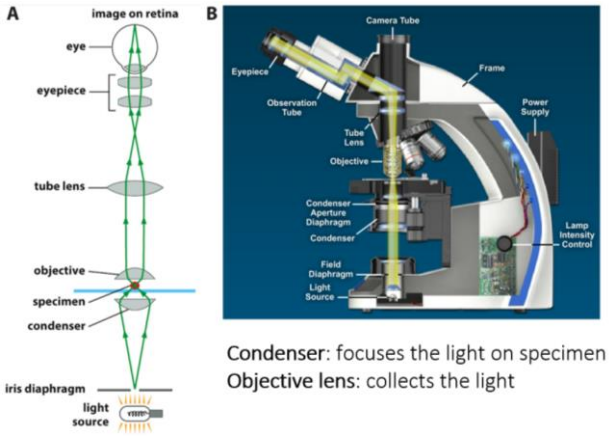
B) distance object-lense > distance focal point-lense
→ Image behind the lense
→Real image

C) Microscopy; Combination of both (A and B)
→objectiv: (as in B) makes real Image
→ ocular: (as in A)
→magnifications are **multiplicative**

- **focal length** (Brennweite): describes the lenses capability to focus light → the smaller the focal length, the stronger the lens
- focus only if the specimen is located at a very specific distance in front of the objective lens
- **depth of field**: distance range in which an object is seen as a sharp image → depends on the magnification
  - the higher the magnification, the narrower the depth of field
  - range in which objects appear as sharp image is usually very narrow

### MODERN MICROSCOPES:

- several lenses are arranged behind each other →complex setup (up to 2000-fold magnification)



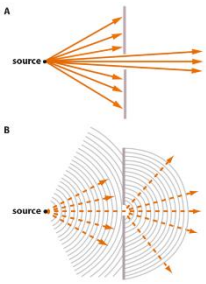

**Condenser**: focuses the light on specimen
**Objective lens**: collects the light

- The overall magnification of a microscope is the product (not the sum) of the magnifications of the objective lens and the eye piece

## DIFFRACTION OF LIGHT LIMITS THE RESOLUTION

Limitation of light microscopy:
A given type of radiation cannot be used to probe structural details much smaller than its own wavelength
→ visible light has wavelengths of several hundred nm
→ we cannot resolve object smaller than 400nm

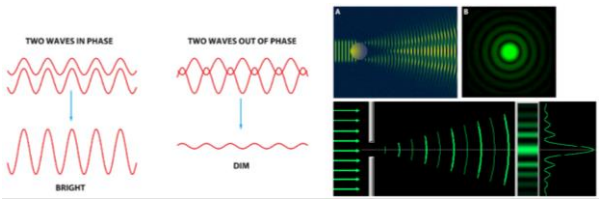
Why?




**diffraction (Beugung):** If wave is hindered in its propagation, it does not simply move just in the direction determined by ray geometry (as a particle – *picture A*)
→ there is a complex wave-propagation (B) pattern outside the geometrical limits of the beam path

- diffraction depends on the size of the obstacle or hole relatively to the wavelenght
  - same size: Wave is diffracted radially (strong)
  - obstacle > wavelenght: little diffraction

- during diffraction, waves **change their path** and therefore **change the phase**
  - waves with the same phase **interfere constructively**
  - out of phase: waves **interfere destructively**
  - certain regions: resulting intensity will be zero → destructive interference of the waves




- object diffracts the light and the waves produce complex interference effects (A)→ Image in a microscopy of such an object will be a **blurred disc** (B)
- the closer two object, the more difficult to distinguish them → at certain distance they look like **1 joint object**

### RESOLUTION:

Minimum distance between two points where they can be still distinguished as separate → **max resolution= λ/2**
  - best resolution possible is around 200nm → resolution limit resulting from the **wave nature of light**
  - **is limited by the wavelength**
  - **depends on wavelength and properties of instrument**

The theoretically achievable resolution of an optical microscope can be **almost doubled**, simply by **using violet light** instead of deep red light → The achievable resolution **increases proportionally with the "shortness" of the wavelength**

- No improvement of microscope or lenses can overcome principal limitation of diffraction (but super-resolution microscopy have higher resolution than ~200 nm)

**Before**: 100x objective lens and a 10x eye piece and you know that this set up will give a diffraction-limited image →you cannot see if feature consists of two small, really close or just of one
**NEW**: You change eye piece to 20x magnification

- New eye piece will **double the magnification** but **won't improve its resolution**
- In a microscope that operates at diffraction limited resolution, a further magnification of the image **does not provide more image** detail → you can't see clearer!
- image seen will be twice as large but will appear "blurry"
- you see less bright than with the old eye piece

→ same amount of light is now spread over a larger area of the viewed image → **A doubling of the magnification will result in a reduction in light intensity** by a factor of 2x2=4

## ELECTRON MICROSCOPY

- Electrons
  - Properties as particle and as wave (but have a mass)
  - Have a smaller wavelength than visible light (1nm)
  - Can be maniplated by lenses
  - Cannot travel with c
  - carry charge and can be manipulated by electromagnetic field
  - Wenn Geschwindigkeit eines Elektrons zunimmt, nimmt seine Wellenlänge ab & seine Energie steigt
- Electron microscopy
  - Wavelenght below 1nm → atomic resolution
  - source of radiation: beam of electrons (not light)
  - electron beam is focused by electromagnets (not by glass lenses)

  -) extensive and demanding sample preparation
- information is contained in the **transmission and scattering pattern** of the electron beam (interaction with nuclei of the atoms or orbiting electrons)
- lenses in electron microscopes are rather imperfect → even the best electron microscopes cannot approach the theoretical limit set by the wavelength of the electrons

## SAMPLE PREPARATION IN ELECTRON MICROSCOPY

### THREE KEY CHALLENGES

- Biological samples…
  - are typically in **aqueous solution** → prevents direct exposure to **vacuum**
  - have **low intrinsic contrast**
  - are **sensitive** to the **radiation** of the electron beam

### APPROACHES

| **Negative staining** | | **cryo-electron microscopy (cryo-EM)** |
|---|---|---|
| sample is embedded in low concentrations of salts of heavy atoms (uranium)<br>→ solution is dried out<br>→ negative can be analysed in el. microscope | | sample is quickly frozen to very low temperatures<br>→ supercooled water is rigid but not organized as crystals<br>→Analyse without vaporizing the water<br>→can see the interior |
| +) provides **high contrast**<br>+) **stable to radiation** | Pros | +) no artefacts<br>+) very high resolutions |
| -) **resolution is limited** to the grains of salt<br>-) frequent **artefacts** due to collapsing structures | Contras | -) low contrast<br>-) more sensitive to radiation → samples have to be treated with less electrons<br>→ worse signal-to-noise ratio → thousands of images have to be combined |

## DIFFERENT METHODS:

- **transmission electron microscopy (TEM)**
- **scanning electron microscopy (SEM)**

produce vastly different kinds of images and have consequently distinct applications

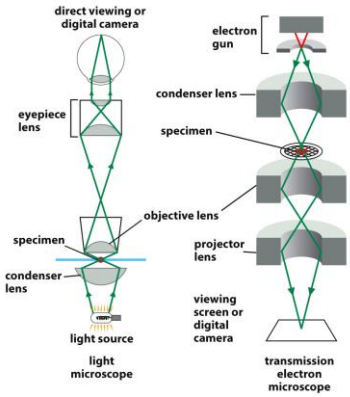
## TRANSMISSION ELECTRON MICROSCOPY (TEM)

- uses high-energy electrons (go through specimen)
- produces 2D projections of thinly sliced samples
- measures the **density** of a sample
- extremely high resolutions up to 0.05nm (atomic level)

limitation:

- **radiation damage arising** in biological samples → limit resolution to **~1nm**
- specimen has to be placed in a vacuum

- measures the **share of the electron beam** that moves through the sample unscattered and thus produces a negative image of the density of the sample
- *misst den Anteil des Elektronenstrahls, der sich ungestreut durch die Probe bewegt und so ein negatives Bild der Dichte der Probe erzeugt*




*electron gun analogue to light source*
*magnetic coils analogue to glass lenses*

- designed similarly to light microscopes –flipped by 180°
- lenses: electromagnetic coils (not glass)
- detector catches electrons that pass through a sample

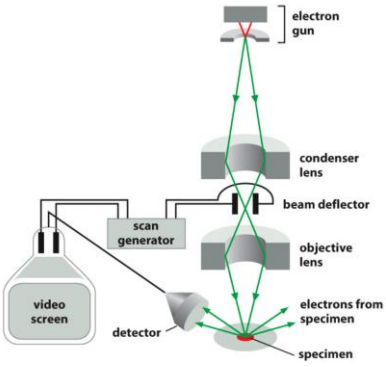
## SCANNING ELECTRON MICROSCOPY (SEM)

- uses a focused beam of **low-energy electrons** (electrons don't go through the sample → just scatter surface)
- scan and reconstruct the **surface** of a bulky object
- provides extremely **high depth of field** (3D-object)
- typically used to **study larger structures** (entire cells or small organisms) rather than subcellular structures
- (Recent developments in electron source allow SEM with high resolution and thus the analysis of small subcellular structures)

limitation:

- only shows specimen's **surface** topography
- → is <u>not</u> useful for examining internal structures
- has a **lower resolution** as TEM → **10nm** (3nm-20nm)

- microscopes is a bit smaller and cheaper than TEM
- detector measures quantity of electrons scattered or emitted



# VORLESUNG ERGÄNZUNGEN

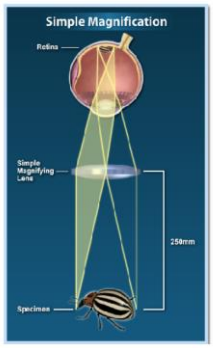
## HISTORY

greek "*micros*" = small – "*skopein*" = to look/see

First Microscopes were build

- in 1590 by Hans & Zacharias Janssen (Holland)
- in 1673 Antioni van Leeuwenhoek reached magnifications of over 200x fold

## MAGNIFICATION = VERGRÖSSERUNG

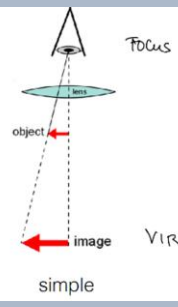



Low Magnification

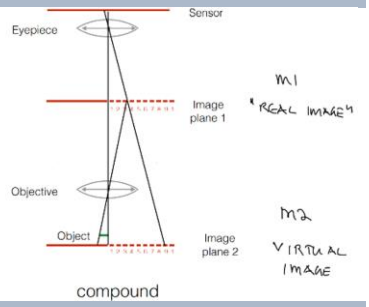

High Magnification

- Magnification does not equal resolution (=Auflösung)!

## FROM SIMPLE TO COMPOUND MICROSCOPES

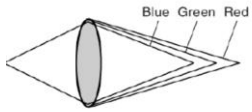



simple

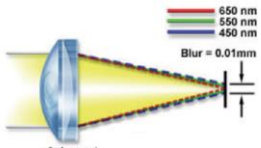



compound

## OBJECTIVES AVE ABERRATIONS




Chromatic aberration



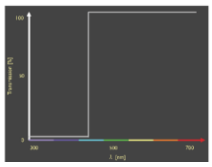

Wavelengths of light bend in different directions (different angles), so that in the end, components of white light have slightly different focal points, resulting in a blur

**Mirrors** (reflect) in addition **with filters** (separate) can select certain wavelengths that are transmitted, whereas the others are reflected:

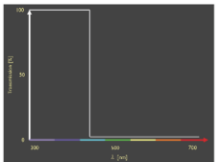
**Short** pass filter: transmits wavelengths **below** a certain point

**Long** pass filter: transmits wavelengths **above** a certain point

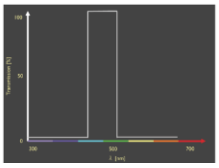
Long pass filter



Short pass filter

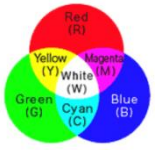


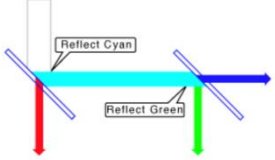
Band pass filter



**Dichroic filters** select for **certain colours** (**RGB-spectre)**

By using chromatic lenses (additional glass piece), the effect can be corrected




QUESTION:

You image cells for one hour either in red, green or blue light. While the cells imaged in **red or green are still alive** at the end of the experiment, cells illuminated with the **blue light are dead after one hour**.

WHY?

1. Because blue light is more refracted than red or green light you have to place the cells closer to the light source to get them in focus. This leads to higher irradiation and phototoxicity
   *WRONG⇒ considering the small distances we are talking about, changes do not really influence the samples*
2. Blue light has higher energy than green and red and is therefore more toxic for cells (true/false)
   *TRUE ⇒ blue is indeed higher in energy (shorter wavelength) and therefore acts more aggressive*

# IMAGING CELLS AND TISSUES AND LIVE IMAGING

## CONFOCAL MICROSCOPY

### WIDE FIELD LIGHT MICROSCOPY

Limitation of conventional light microscopy

- image resolution of fluorescence microscopy is limited by the diffraction of light to approximately 200nm
- **thickness** of a sample → light will be reflected
  - must be thinner than 20um
  - thicker samples produced a **blurred image** due to the **high emission of out-of-focus light**
  - the thinner a sample, the sharper the observed image → less unfocused light from other regions blurs it

2 MAIN STRATEGIES TO REDUCE OUT-OF-FOCUS BLUR:

1. optical – confocal microscopy
2. computational image-processing = deconvolution

IMPROVE CLASSICAL FLUORESCENCE MICROSCOPY:

- Both, **deconvolution** and **confocal microscopy** try to overcome limitations of classical light microscopy
  - Limitation through **out-of-focus blur** → Both
  - Limitation **through diffraction** → Neither
  - Limitation through **need for thin sample** → Confocal

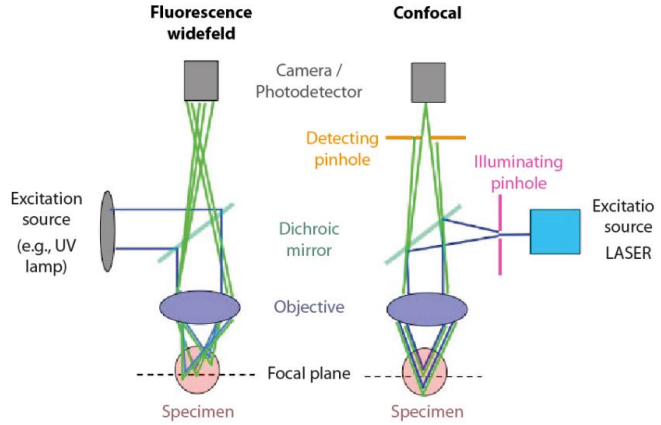
### CONFOCAL MICROSCOPY

→**focus** on a chosen spot in a thick sample and **reject the light** that comes from regions above or below this region
→ Point-illumination and point detection
→optical section

confocal microscopy **restricts the path of the light before it is measured**

- manipulation takes places in two ways
  - light does not illuminate the complete sample but just a single point → laser light passes through a pinhole (illuminating pinhole)
  - light that is emitted from the sample passes through another pinhole before being collected by a detector (detecting pinhole)
- detector pinhole is placed ***con focal*** (= same focus) with the illuminating pinhole
  - only light that is emitted from the focal point passes through the detector pinhole and enters the detector
  - light that is emitted from regions out of focus is excluded from detector →light cannot pass 2. Pinhole

high-resolution image of a single point is made by the *con focal* positioning of two pinholes



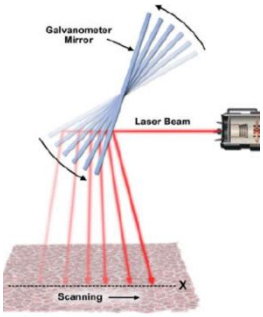

LIGHT SOURCE:

- not a normal light source but a **Laser** → light is focused into a very small spot of the sample

SCANNING:

- A lot of small images are collected to build the entire image of the sample (Man muss das Bild aus vielen Einzelnbilder zusammensetzen!)
- the detector is attached to a computer that builds up the image, one subimage at a time
- Scanning methods:
  - move the **specimen** → takes a very long time and it's necessary to drive the stage with great precision
  - move the **light source** → very fast and ideal for capturing very small shapes → more often used!

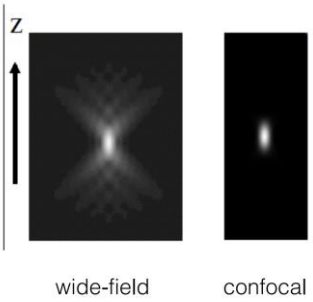
LASER SCANNING




- To move light stepwise within the focal plane the laser beam is **deflected by an oscillating mirror**
- Images is illuminated point by point by scanning the laser beam across the field of view
- light is detected with a point detector (PMT) (photo multiplier tube)
- It's possible to scan from top to bottom and examine the object slice per slice → **three-dimensional image of the object**

SUMMARY:

- solves the problem of out-of-focus light (Lase + Pinholes)
- allows a reconstruction of a **three-dimensional image**
- does exclude background fluorescence from other planes and thus increases contrast → the image looks sharper
- does not increase resolution
- roughly one eukaryotic cell can be captured per image
- (At a 1000-fold magnification, an area of 20x20 µm is detected)
- the **smaller the pinhole the sharper**, but at the same time the **darker the resulting** image, as less photons hit the specimen → **sharp-vs.-bright compromise**

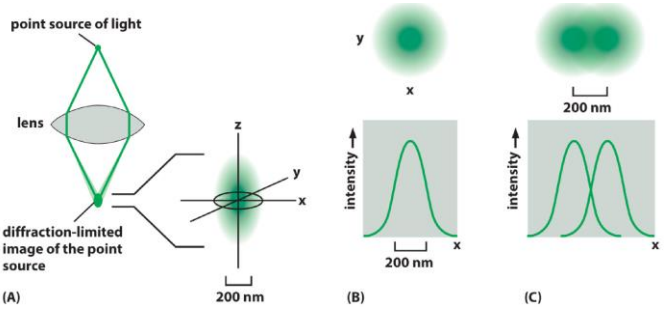


## IMAGE DECONVOLUTION

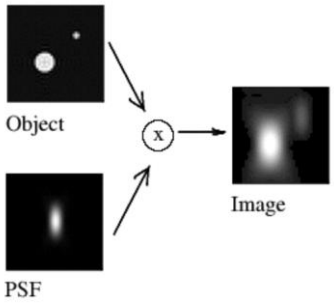
Helps to convert (deconvolut) blurred image into clean image

REMEMBER: **wave-like nature of light**

- **interaction of light with an object leads to diffraction** of the light wave resulting in a **loss of phase coherence**
  → Waves that are out of phase interfere destructively, waves that are in the same phase interfere constructively
  → in certain regions the resulting intensity will be zero
  → single point seen in a microscope appears as a blurred disc
- **diffraction effects** lead to no clear point but a blurred one in all dimensions




- **point spread function**: Gaussian distribution to describe the blurred image of a point seen in a microscope
- computational deconvolution "subtracts" the blurring effect that a microscope has on an image



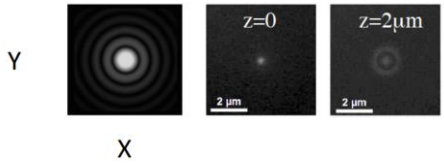

- Back calculation (=deconvolve) to the real object
- one has to know the blurring effect of the microscope
  → point spread function measured by computer (first!)
- deconvoluted images are still constrained by the diffraction limit
  - helps to clear images form **out-of-focus light**
  - it **cannot improve the resolution** of the image

# ERGÄNZUNGEN DER VORLESUNG

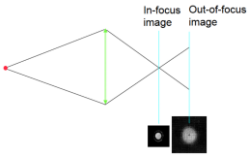
## DIFFRACTION AND OUT-OF-FOCUS LIGHT

→ blur images in microscopy

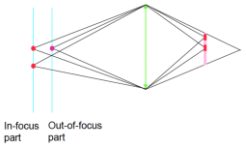
DIFFRACTION:



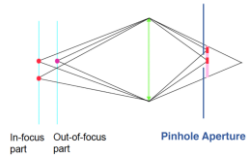

OUT-OF-FOCUS:




3D: In a 3D object there is cross-talk between in and out-of-focus parts

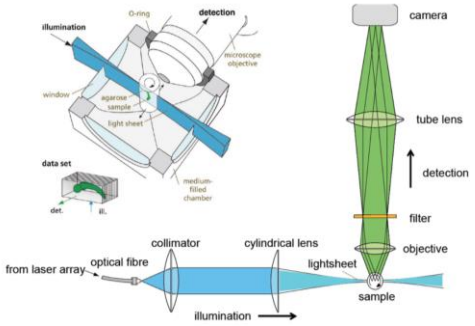



CONFOCAL MICROSCOPY:

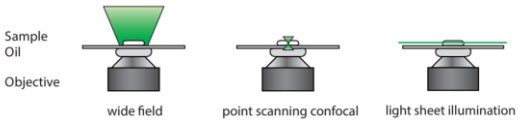



## LIGHT SHEET ILLUMINATION

- Excitation light is focussed to a sheet of light that illuminates only the focal plane of the detection optics




## COMPARISON

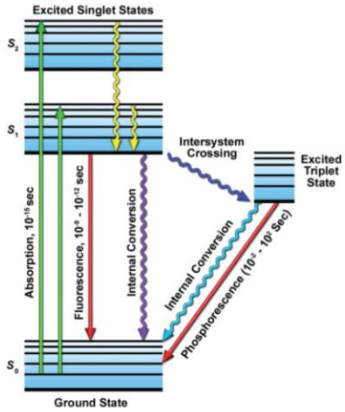



NA~1.0 -1.4 NA = numerical aperture [light gathering ability of objective]

| | Wide field | Laser scanning confocal | Light sheet |
|---|---|---|---|
| xy resolution | 0.61λ/NA | 0.4λ/NA | 0.4λ/NA |
| z resolution | Low | Up to 500 nm; variable by varying pinhole diameter | High |
| brightness | very bright (200x) | dim (1x) | bright |
| speed | fast | slow | fast |
| thick samples/high background | blurred | good sectioning up to 100 um into sample | excellent sectioning |
| photodamage | low | high | very low |
| other | Easy-to-use | Zoom | Complex light path and set-up. Limited commercial availability |

## FLUORESCENCE MICROSCOPY

**Fluorescence** = electrons return **to a ground state** emitting lower-energy (large wavelength) photons




*Jablonski Energy Diagram*

1. Electrons are **excited** by **absorption of a photon**
2. **Fluorescence** is observed, when it returns back to its ground state. Some of the initial energy provided by the photon is **lost as heat**, wherefore the emitted photon contains fewer energy (**quantum efficiency**)
3. Cycle can be repeated until **photobleaching**: ability of fluorescence of a molecule is limited, because exposure to light stresses them, resulting in photochemical damage

→ reducing **light-intensity** or **exposure-time** reduce damage

### LIGHT DAMAGE:

- Each fluorescent molecule be photo-excited several times BUT it's possible, that absorption of a photon leads to a **photochemical reaction of the chromophore** → In this relatively rare case, the fluorophore's **chemical structure is altered** →results in **permanent change of its properties**
- The energy carried by photons of visible light **approaches the energy needed to break covalent chemical bonds** → Every time a fluorophore is photoexcited, there is a small (typically >> 1 %) but finite risk that this excitation triggers a photochemical reaction, which may destroy the fluorophore molecule

### $S_0$-STATE:

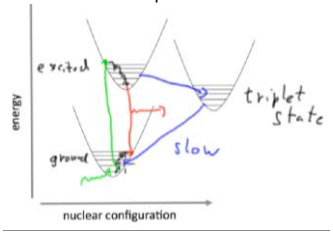
- The ground state $S_0$ is the **preferred** state → after excitation, molecules rapidly **return** very quickly to the $S_0$ state
- fluorescence microscopy → uses relatively **modest light intensities** → the vast majority of fluorophore molecules in the sample will be in $S_0$ state ready to absorb a photon

## TRIPLET STATE:

Additional type of electronic state

Likelihood relative low, but has a really long lifetime

- no fluorescent signal (Circulation is interrupted) → dark form of fluorophore

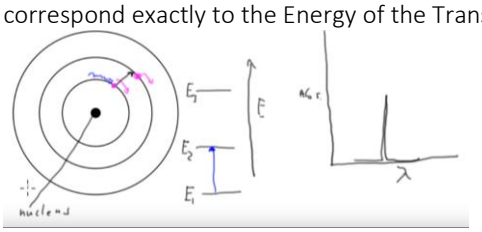



## ABSORBANCE/EXCICATION AND EMISSION

**EXCITATION** spectrum: range of wavelengths that are able to **excite** an electron from ground state to a higher energy level

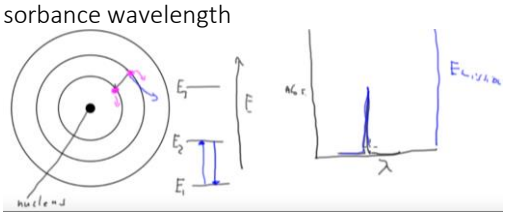
Photon must have the exact energy that is needed to push the electron in a higher orbital (energy level)

- Absorbance peak → Energy of the photon (wavelenght) correspond exactly to the Energy of the Transmission



**EMISSION** spectrum: range of certain wavelengths that are **emitted** when an electron returns to its ground state

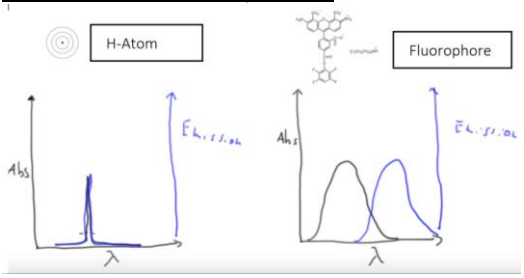
- Electron lowers an orbital → emission of energy in form of a photon → H-atom has the same emission and absorbance wavelength



- spectras are **almost equal in shape, but have different peaks** → emission spectrum more right (longer wavelengths, fewer energy) →this phenomena = **stokes shift**

## H-ATOM VS FLUOROPHORE

**Absorbance-Emission spectrum**:




**H-Atom (left):**

emission spectrum = absorbance spectrum

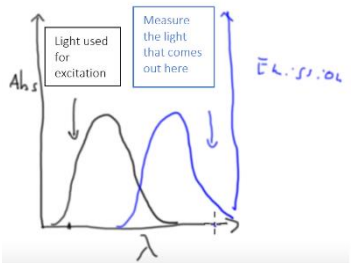
→ the emitted light has the same wavelength as the absorbed light (not useful because we cannot distinguish between light that goes in and comes out)

**Fluorophore (right):**

1. **absorbance spectrum ≠ emission spectrum**
2. **broader peak**

## 1. SHIFT IN THE WAVELENGTH

→ we can use a different wavelength for excitation (smaller wl) so we can measure the emission by a higher wavelength → we can well separate the light that goes in with the one the comes out

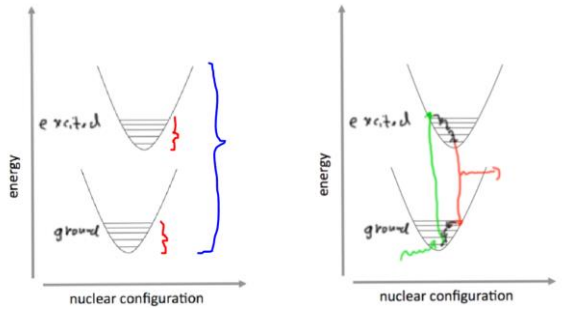



***Why is the Emission shifted to longer wavelength (lover energy)?***

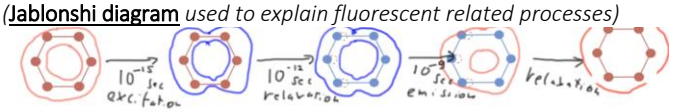
complex molecules have also nuclear configuration (= arrangement of the course of the atoms in the molecule) → can change in the process of excitation and de-excitation → leady to conformational changes

**energy levels of a complex molecule (left):**

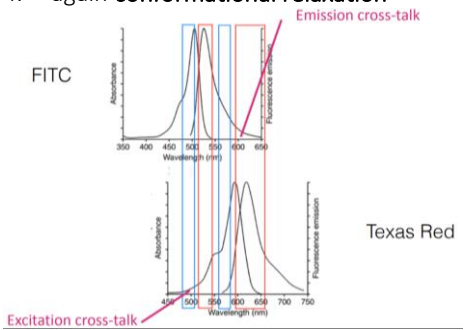
- electronic energy levels (excited/ ground state)
- vibrational energy levels; different vibrational states




(Jablonshi diagram *used to explain fluorescent related processes)*



1. photon is coming in and excites the electron → change of the electronic distribution (nuclei distribution stays the same but electron distribution changes→large mismatch)
2. this leads to a **relaxation process** of the configuration of the nuclei and the surrounding solvent molecules → relax in the levels of vibration to the ground → new atom configuration (equilibrium with surrounding electron cloud) → we are electronically in the excited state but configurationally in the relaxed state
3. emitting a photon (= fluorescence) → energy of emission will actually be lower cause we lost energy in relaxation and jump into a higher configuration again (right-shifted wavelength)
4. again **conformational relaxation**




- Spectra are **influenced by environment** (pH, ionic strength, $O_2$ concentration etc.) → can be manipulated
- **Filters** are used to **select non-overlapping regions** for excitation and detection
- **excitation filter**: cuts of UV light, which, due to its high photon energies, is liable to cause photo-damage of the sample, but still allows excitation of the chromophore

## 2. BROADNESS (=BREITE)

- The absorption and emission spectra of the fluorescent molecules used in microscopy **tend to be rather broad**

**Broad** (= breit): a lot of different wavelength are able to excite the electronic transition in our fluorophore

### WHY BROAD?

Fluorophore molecule is **surrounded by solvent molecules** → these bump into fluorophore molecules → thermal energy causes slightly different conformation in different Fluorophore molecules → **different conformations have different energy levels** and therefore correspond to a different region in absorbance-spectrum

- Fluorophore have **different conformations which will absorb energy at slightly different wavelength** → add up all individual absorbance spectra of different conformations leads to the broad absorbance spectrum!
- **Due to intramolecular motions**, different fluorophore molecules will undergo electronic transition (or emission) of photo-excitation in **different conformations** → the exact energy of the transition will **depend on the conformation of the molecule** at the time transition takes place
- electron **does not always emit the same amount of energy** when it falls back into the ground state → Depending on the exact conformation of the fluorophore the energy and wavelength of the photon may vary
- electronic transitions from excited (singlet) state to triplet state (and reverse) are **not accompanied by the emission of a photon** → don't contribute to the emission spectrum
- Emission spectra aren't broad because the excitation light contains many different wavelength → WRONG!

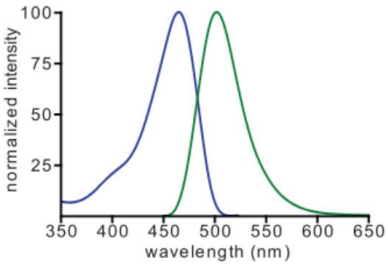
## QUANTUM YIELD

- The photon energy absorbed by a fluorophore **is NOT always re-emitted as a photon**
  - energy of the absorbed photon may be released through a **series of mechanical collisions** with surrounding molecules
  - **internal-conversion**: absorbed energy is released as **heat**

  → fluorophore always returns to its ground state and is ready for another round of photoexcitation
- **quantum yield** = the fraction of excitation photons that results in the re-emission of fluorescent photon
  - quantum yield of 1; each excitation photon leads to the emission of a fluorescent photon
  - fluorescent proteins often have a quantum yield of ~ 0.6 → 40% of the excitation photons absorbed by these fluorophores do not lead to the re-emission of a fluorescent photon.
- Emission des Fluoreszenz-Photons die Fluorophor-Moleküle molekulare Kollisionen durchlaufen, die die angeregten Fluorophore über alle möglichen Konformationszustände wieder ins Gleichgewicht bringen

AUFGABE:

absorption and emission spectrum of a fluorophor:




- filter should not let through any **excitation light** but still cover a large portion of emission spectrum
  → for fluorescence microscopy with this fluorophore, a filter pair consisting of a **400-500 nm bandpass filter for excitation** and a **510-600 nm bandpass filter for detection** would be ideal
- The emitted photons will have **longer wavelength** than the absorbed photons → After photon excitation the fluorophore relaxes conformational → some energy of the absorbed photon is lost prior to its re-emission
  Fluorophors macht nach Photonenaufnahme eine konformative Relaxation, wodurch ein Teil der Energie des absorbierten Photons verloren geht, bevor dieses wieder emittiert werden kann
  → The **blue** spectrum is the excitation spectrum and **green** the emission spectrum (the inverse is not possible)
- **Stokes shift (wavelength shift)** = wavelength difference between the peak of the absorption and the peak of the emission spectrum → here it is 45 nm
- When this fluorophore is illuminated first with narrow wavelength light at exactly 400 nm and later with broadband light from 450-475 nm, the emission spectrum of the fluorophore will have exactly the same shape (But its intensity may be different) → YES; Light of any wavelength (as long as it falls within the absorption spectrum of the fluorophore) will result in the **excitation of the chromophore** and will result in **the emission of photons across the entire emission spectrum** → between the absorption of the excitation photon and the emission of the fluorescence photon, the fluorophore **molecules will undergo molecular collisions** that re-equilibrate the excited fluorophores across all possible conformational states

Licht jeder Wellenlänge (solange es in das Ab-Sorptionsspektrum des Fluorophors fällt) führt zur Anregung des Chromophors und zur **Emission von Photonen über das gesamte Emissionsspektrum**. Der Grund dafür ist, dass zwischen der Absorption des Excitations-Photons und der

# IMAGING DYNAMIC PROCESSES

## DYNAMIC CELLS

- cells are not static → are highly dynamic!
- Movement of cellular macromolecules – 2 categories:
  - **Diffusion** → Brownian/ thermal motion of molecules
  - **Active transport** → driven by the hydrolysis of energy-rich molecules such as ATP (speed up movement and restrict direction of molecules)
- Additionally, molecular movement is influenced by the high concentration of biomolecules in the cell

→ CELLS ARE INCREDIBLY **CROWDED**!

- **slow down diffusion**
- **speed up interactions** because there is less 'free space' and molecules are more likely to bump into each other

## MEASURE DYNAMIC CELLS

techniques to measure dynamic events/ diffusion in cells:

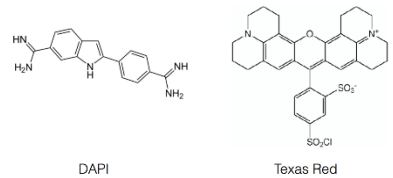
- single particle microscopy
- **FRAP** (photobleaching → observe motion, diffusion)
- **FRET** (transfer of energy → interaction an conformational changes)

## FLUORESCENT PROTEINS

- can be used to visualize dynamic processes
- they specifically tag proteins and make their location visible by **excitation with specific wavelengths** in fluorescence microscopy
- Most organic fluorescent molecules are **chemically synthesized** and then **introduced into the cell**
- **green fluorescent protein (GFP)** is encoded by a single gene, which can be introduced into the cell's genome → place under the same transcriptional control as a gene of interest → fusion protein can still fold and function normally but GFP fluorescence allows the observation

### FLUORESZENT MOLECULES

- **Organic** molecules (contain **large conjugated systems** on which electron excitation relays):



DAPI Texas Red

- **Proteins**: GFP (green fluorescent protein)

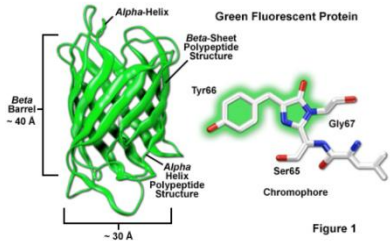



Figure 1

## VISUALIZE MOTION WITHIN A CELL

### SIMPLEST WAY:

- take several images at different time points
  - can see if and how molecules are moving around
  - can observe if they move randomly, by diffusion, or if it is a directed movement
  - cannot measure the **underlying kinetics** of movement

## PHOTOBLEACHING

- measures **diffusion**, **quantify the motion** of biomolecules in cells and **protein binding and dissociation rates**
- it's a **chemical modification of the fluorophore molecule** → induced by **exposure to light of high intensity** → resulting in a **loss of fluorescence**

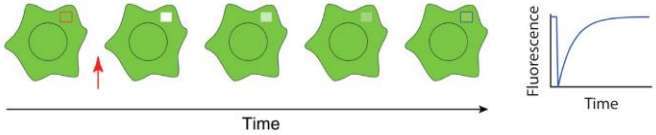
# FRAP

## FLUORESCENCE RECOVERY AFTER PHOTOBLEACHING

- with FRAP you can **study the motion of molecules**
- phenomenon of photobleaching is used to observe the motion of molecules
- helps to determine if objects move due to **random brownian motion** or due to an **active transport process**
- isn't limited to measurement of diffusion processes
- gives information about **kinetic parameters** of a protein
  - diffusion coefficients
  - active transport rates
  - binding and dissociation rates from other proteins

### MECHANISM:

1. strong laser beam is focused to photobleach fluorescent molecules in a certain part of the cell
2. observe the recovery of the fluorescence → unbleached fluorophores from other parts diffuse to bleached area




- kinetic coefficients can be calculated (**diffusion coefficient)** by plotting the mean fluorescence in the bleached area versus the time after photobleaching
- REMEMBER: The fluorophore molecules that are bleached in a FRAP experiment **are permanently destroyed** and do not recover → The observed recovery of fluorescence is due to fresh, unbleached fluorescent molecules (from parts of the cell, that were not exposed to the bleaching laser light) moving into the bleached area

### PROs:

- photobleaching process doesn't change the protein concentration in the bleached area → kinetics of the recovery represent the normal dynamic behavior of the observed protein
- can measure diffusion but also protein binding and dissociation rates

### MEASURING DIFFUSION:

- the more mobile the protein, the faster the recovery
- no recovery if it's immobile (attached to plasma membrane, strict located in a certain organelle etc.)
- fluorescence does not reach its initial level: just a fraction of proteins is immobile

### MEASURING PROTEIN BINDING AND DISSOCIATION:

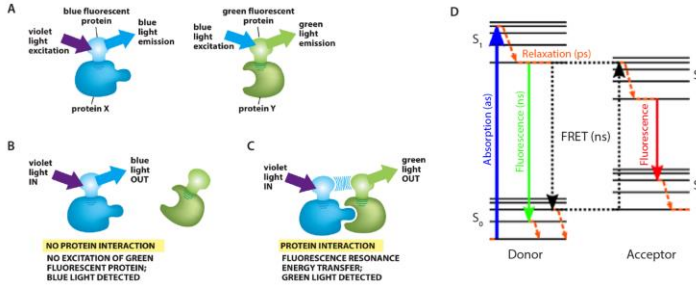
- if the fluorescently tagged protein binds to **static receptors** in the bleached area, the recovery of the fluorescence reveals information of the association and **dissociation constants**
  - slow recovery: protein is tightly bound and disassociates slowly
  - fast recovery: recycling of the bound protein is rapid

# FRET
## FLUORESCENCE RESONANCE ENERGY TRANSFER

- In FRET, the proximity (Lage/ Abstand) between two fluorophores leads to a transfer of energy and therefore to a change of the observed emission wavelength
- Monitoring **molecular interactions by FRET**
- can detect that two molecules are within several nm
- FRET technique has a wide range of application and can be used to…
  - study certain protein-protein interactions
  - monitor the concentration of a molecule in a cell
  - measuring structures, conformational changes, interactions between molecules, and biochemical events

### HOW DOES IT WORKS?

- **energy shift between excitation and emission of a fluorophore** is the key to how FRET works
- two molecules are labeled with different fluorophores
  - (D) emission spectrum of fluorophores one (the donor) overlaps with the excitation spectrum of the other one (the acceptor)




- (A) Protein X is coupled to a blue fluorescent protein
  → exited by violet light (370-440nm)
  → emits blue light (440-480nm)
- (B) Protein Y is coupled to a green fluorescent protein
  → exited by blue light (440-480nm)
  → emits green light (510nm)
- Illumination sample **only with violet light:**
  - (C) Protein X and Y don't interact: blue light is emitted
  - (D) Protein X and Y interact: green light is emitted
    → donor fluorophore transfer energy from the absorbed light directly to the acceptor fluorophore
    → **FRET occurred**

- transfer of energy between the two fluorophores **happens by resonance** and is therefore **non-radiative** → no energy is lost in the transfer and no photon emission

## FRET EFFICIENCY

- used to study the structure and interactions of proteins
- criteria that must be fulfilled for FRET to happen
  - overlapping excitation and emission spectra
  - fluorophores must be close to each other (1-10 nm)
  - lifetime of donor molecule must be long enough to allow the energy transfer to occur

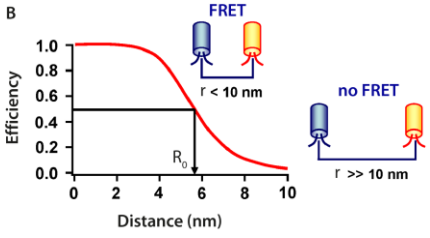
Efficiency of the FRET process ($E_{FRET}$):

- $E_{FRET}$ = transfer efficiency lies between 0 - 1
  - 1 = entire energy transfered)

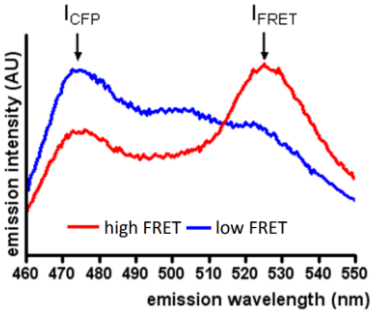
A

$$E_{FRET} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}$$

B




r = **donor-aceptor-distance**
$R_0$= **Förster distance**= Distance of a FRET pair where transfer efficency is 50% → varies for ech donor-aceptor pair (often 2-8nm)



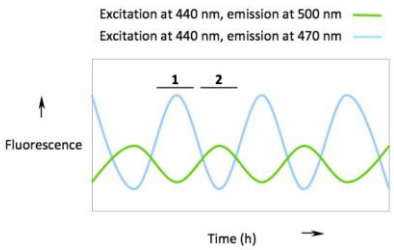

- FRET efficiency has a **strong distance-dependence**
- distance over which energy can be transferred **strongly depends** on the used **fluorophore pair**
  - ideal pair has a strong overlap of the donor emission and the acceptor excitation spectra
  - higher overlap between spectra => larger Förster distance values → energy transfer can happen at a greater distance

### EXAMPLE → *ran-GTP; siehe Vorbereitung p.6*

Imagine a transcription regulatory protein (X) that is known to shuttle back and forth between nucleus and cytosol in an oscillatory pattern. Protein Y is a nuclear protein that can bind to X to create a dimer that binds to DNA. You have fused protein X to green fluorescent protein (GFP) and protein Y to blue fluorescent protein (BFP), and have measured fluorescence resonance energy transfer (FRET) and non-FRET signals in the nucleus at different time points, as indicated in the following simplified plot.

**At which time period (1 or 2) do you think protein X is in the nucleus?**

BFP can be excited at 440 nm, and emits maximally at 470 nm. GFP is excited at 470 nm and emits maximally at 500 nm.

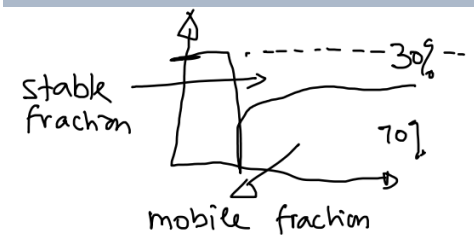



→ Correct answer: 2

The FRET signal in the green curve (negatively correlated with donor fluorescence in the blue curve) is expected to be high when the two proteins can interact in the same compartment. Additionally, the lower-energy wavelength of emission (here, green) can in general only be observed in a FRET experiment when the two fluorophores are close together.

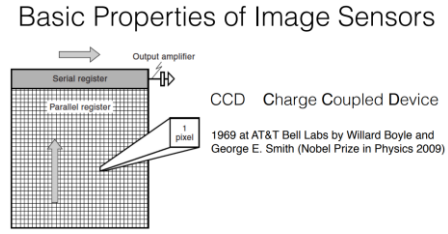
# ERGÄNZUNGEN DER VORLESUNG

## FRAP

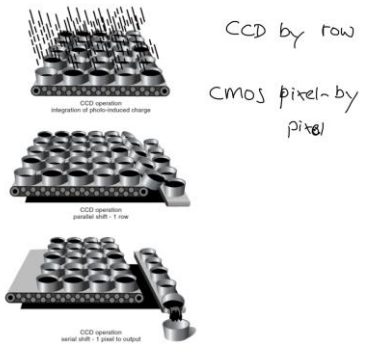


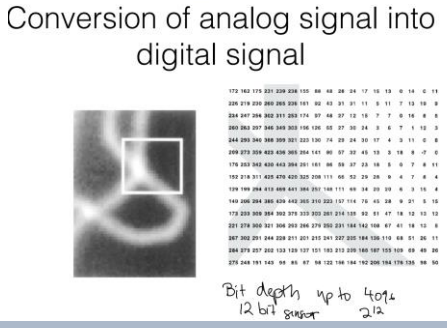

## DETECTOR/ SENSOR


- Digital image sensors convert analog light signal into digital signal

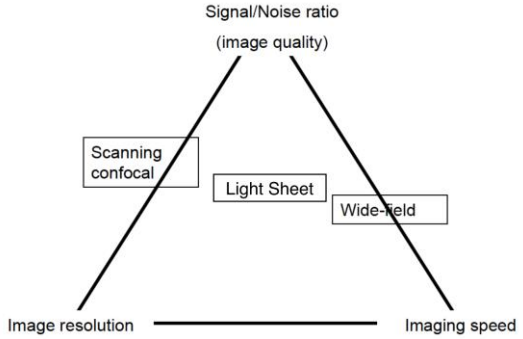



CMOS: Complementary Metal-Oxide-Semiconductor




## TRIANGLE OF COMPROMISSES




## PIXEL SIZE

QUESTION:

***How small do the pixels on a microscopes camera have to be in order to resolve two objects that are 200 nm apart?***

- Nyquist theorem: in order to adequately reproduce a signal, it should be sampled at intervals (i.e. with a pixel size) that **is half as small as the resolution**, at which one wishes to record the image

Das Nyquist-Theorem besagt, dass ein Signal, um es adäquat wiederzugeben, in Abständen (d.h. mit einer Pixelgröße) abgetastet werden sollte, die halb so klein ist wie die Auflösung, mit der man das Bild aufnehmen möchte

- The **magnification of the objective lens is 100x**
- No eyepiece is used

***Calculate the edge length (i.e. size) of a pixel that can achieve this diffraction limited resolution***

SOLUTION:

200nm (apart) * 100 (magnification) = 20000nm = 20um apart
→At 100 x magnification the two objects would be 20 um apart → According to the Nyquist theorem one would need to sample the image with a pixel size that is < 1/2 this distance
This corresponds to a pixel size of **10um** or smaller

# SUPERRESOLUTION MICROSCOPY

- all light-microscopy are limited by a physical restriction as to the size of structures possible to resolve
  - no resolution better than 0.2µm → objects have to be further apart in order to be distinguished
  - because of diffraction → limits resolution → blurred

## SUPERRESOLUTION BY SINGLE-MOLECULE LOCALIZATION

Superresolution techniques allow localization of fluorescent molecules to accuracies much below the diffraction limit of resolution, which is about 200nm

## PHOTOACTIVATED LOCALIZATION MICROSCOPY

= PALM

- overcomes diffraction barrier (0.2µm) of light microscope
- Instead of exciting and detecting all fluorophores at once, we **look at them individually** → repeat many times and then combine/reconstruct "superresolution" image
- fluorophores appear as overlapping blurry discs but by **exciting so few fluorophores at once** it's very unlikely that two fluorophores closer than 0.2µm are excited

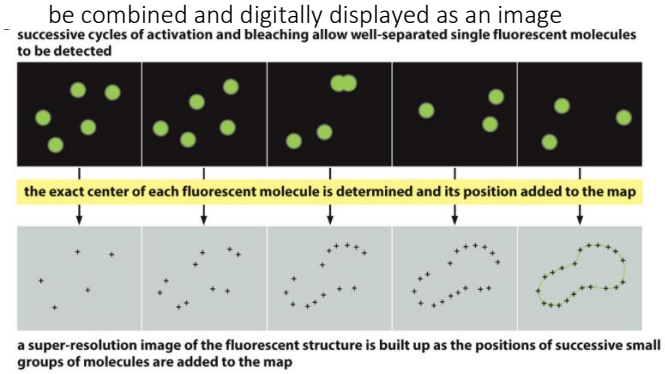
### PA-GFP

- special photoactivatable fluorophores play a crucial role, because each cycle is only performed on a small subset of molecules that is activated by a high-energy laser
- photoactivatable green fluorescent protein
- doesn't fluoresce until activation by illumination at a certain high-energy wavelength
- has a mutation in one amino acid (vs. wild-type GFP)
- protein shows almost no fluorescence at its normal excitation wavelength when no photoactivation at 413nm has occurred prior to excitation
  - **activation** wavelength at 413nm *(NEW!)*
  - **excitation** wavelength at 488nm
  - **emission** wavelength at 515nm

### METHOD

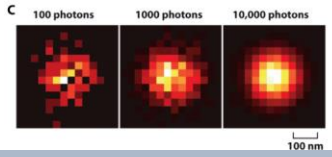
- repeated cycles of fluorophore activation, localization, and bleaching result in nanometer-resolution images

1. illuminate sample by a **laser at activation wavelength** with low enough intensity → only few random fluorophores are activated while the remaining ones stay inactive
2. excite activated fluorophores by another laser at excitation wavelength and detect their fluorescence
3. photoactivation, excitation, single-molecule localization, and bleaching are repeated until all fluorophores have been imaged → the exact coordinates of the full set can be combined and digitally displayed as an image



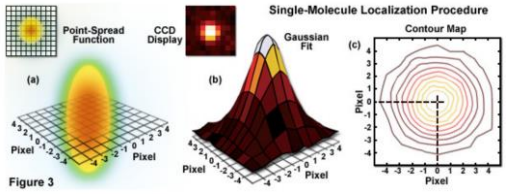

### BLEACHING AND ACCURACY:

- Frequently used fluorophores are bleached after a few 100 - 1000 excitation cycles and **cannot longer be excited**
- After every excitation the fluorophore **emits one photon**
- **a few thousand photons** are required to determine the position of the fluorophore computationally to precision of a few nm
- number of detected photons (= excitation cycles) determine the precision (=accuracy) with which a fluorophore can be localized → Determination of the exact mathematical centre of the blurred image becomes more accurate the more photons contribute to the finale image
  → **the more photons, the better the resolution**




## POINT SPREAD FUNCTION

- PALM technique doesn't reduce the width of the point spread function (doesn't change the shape of the point spread function) but the **center of the point function is determined**



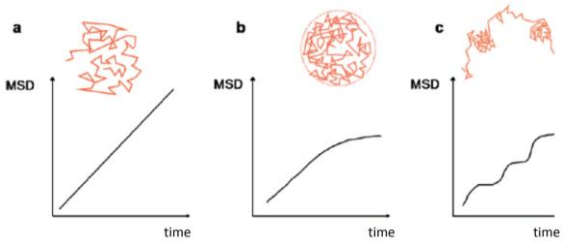

## LIMITATION OF PALM

- light-microscopy: resolution limit because diffraction
- natural limit of PALM:
  - the size of the fluorophore
  - the distance (need to be spaced well apart)
- resolution limit **~ 10nm** (ø distance between macromolecules within cells)
- resolution will typically be limited **by diffusion and exposure times** and not by the effects of diffraction
- **sensitivity to photo-bleaching** limits the resolution
- bleaching of fluorophores as a **stochastic process**
  → accuracy of position of fluorophores differs!
  - **trade-off:** high density of molecules but exclusion of those fluorophores whose exact position cannot be determined with high confidence (but not too many!)
  - the more molecules (higher density) and the exacter their positioning, the higher the resolution
- tracking the position of individual molecules with higher precision **will reduce the length** over which it can be tracked → increase the photons by increasing light intensity → **increases risk for photobleaching** of fluorophore
- fluorescent marker only reports information about **location of a molecule**, but **not its structure or position** relative to the fluorophore
- super-resolution is assembled from multiple consecutive images and those images are recorded at different time points so that **information about the coordination of movement is lost**
- molecules shouldn't move quickly → fitting procedure for molecule position will not work properly→ exposure times of the individual images need to be sufficiently long
- cannot work with unstained samples→ need of spatially well-separated light sources

# VORLESUNG ERGÄNZUNGEN

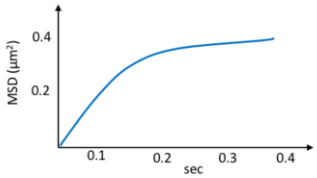
## MEAN SQUARE DISPLACEMENT

Left: free diffusion




AUFGABE:

the plot shows **the square of the average distance** of the object from its starting point as a **function of time**, not the speed with which individual molecules move along on their random path. Therefore, the plot is consistent with a constant speed of the individual molecules



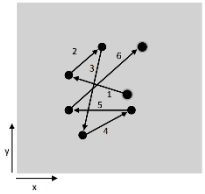

- No free diffusion! → distance from the starting point doesn't continue to increase → From about 0.2 seconds after the tracking started, the distance from the starting point stays approximately constant
- Here the diffusion of the molecules being tracked are confined to volumes with a diameter on the order of 1µm. (Diffusion beschränkt sich auf Volumina mit 1µm Durchmesser)
- Im Anfangsteil der Kurve nimmt die MSD linear mit der Zeit zu, was auf freie Diffusion hinweist, aber ab 0,15 Sekunden steigt der Abstand nicht weiter an. Das bedeutet, dass die Moleküle innerhalb von 0,15-0,2 Sekunden das volle Volumen erforscht haben, zu dem sie Zugang haben. Der Radius dieses Volumens entspricht in etwa der Quadratwurzel dieses MSD

## CROWDING

- **slow down the speed of diffusion**
- **can enhance molecular interactions** → because there is less 'free space' and pairs of molecules are more likely to bump into each other repeatedly

## BROWNIAN MOTION

- Time-average displacement is zero (if no bias or energy input)
- Time-average squared displacement is not zero (unless there is no movement at all)




## EQUILIBRIUM

- At equilibrium there is no net movement of molecules within a cell
- Under equilibrium conditions, individual molecules that are not trapped or attached to larger structures in the cell will move through the cell via diffusion. The exact path of these motions is random

## QUIZ:

The motor protein kinesin carries cargos along microtubles. For each **ATP molecule it hydrolyses it takes an 8 nm step** and under optimal conditions it takes **about 50 steps per second.** You are observing this process with a microscope. The diameter of the image you observe with your microscope's camera corresponds to **0.2 mm of the object** (this is a typical value for a modern microscope with a 100x objective)
How long does a kinesin molecule take to traverse this field of view in units of seconds? 500

$8x10^{-9}m * 50\ s^{-1} = 4x10^{-7}m/s$
Step size * rate constant of stepping = speed of the kinesin molecule
$4x10^{-7}m/s\ /\ 0.2x10^{-3}m = 500\ s$
Speed / distance to be traveled = time required to traverse the field of view

In living cells, soluble molecules, unless they are trapped by or attached to larger structures (e.g. cell membranes or organelles), are in constant motion via diffusion.
If the **distance these molecules diffuse during the exposure time of a microscopic image is large,** relative to the resolution of the image, **the image will be blurred**. To find out, if this effect is a problem in fluorescence microscopy, please calculate the time a protein molecule, such as GFP, takes to diffuse the distance that corresponds to 3 times the diffraction limited resolution of a conventional fluorescence microscope

- diffusion coefficient of GFP in eukaryotic is D= 30 $\mu m^2/s$

**diffraction limited** resolution of a conventional light microscope is ~200 nm, which equals **0.2µm**
**$t = x^2/2D$** (x = distance = 3 x 0.2 = 0.6 µm)
= 0.006s = 6us
6us is shorter than typical exposure time used to record a microscopic image (recording microscopic images: exposure times often 0.02 sec) → freely diffusing GFP molecule will therefore appear blurred → for fluorescent molecules that are not tightly attached to larger structures, the maximally achievable **resolution will typically be limited by diffusion and exposure times and not by the effects of diffraction**

- diffusion coefficient of a ribosome (D = 0.04 µm2/s)

more exposure time than needed → this case the achievable resolution is therefore limited by the diffraction limit

## TIRF – TOTAL-INTERNAL REFLECTION FLUORESCENCE MICROSCOPY

- Spatial restriction reduces background noise
- Allows the detection of single fluorescent molecules

### UNWANTED LIGHT SOURCES

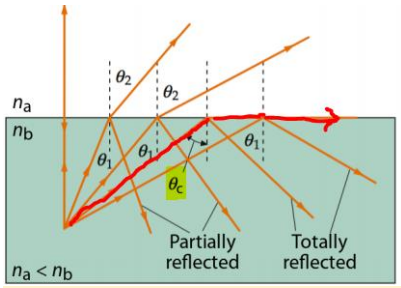
- In ordinary microscopes, it's difficult to detect single fluorophores without detection of background noise
  - light emitted or scattered by out-of-focus molecules confer a strong background to the image
  - additional source of unwanted fluorescent signal due to **autofluorescence** of some structures (mitochondria, chloroplasts, NADH)

➔ all this source of light tends to blur the fluorescence from the particular molecule of interest

### REFRACTION

= light rays change their direction

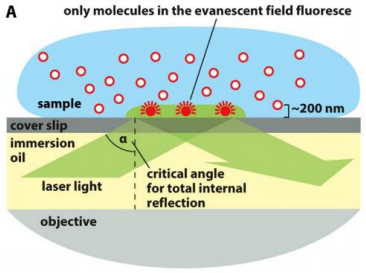
- occurs when light travels from one material into another
- rays are bent **away from the normal** of the surface, if light travels from a dense material (high refraction index) into a less dense material

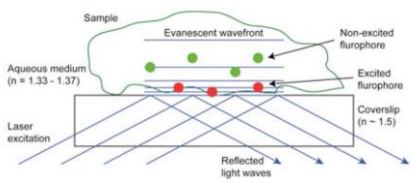



- **critical angle of total internal reflection**: light rays are bent such they **do not leave the dense material** but **travel at its surface** → occurs if the angle between the incident ray and the normal is larger than a critical value
  - light travelling from glass into air: angle $\theta_c = 42°$
- At a larger angle, light will be completely reflected back into the dense material

### TIRF MICROSCOPE

- specimen is illuminated with a laser at an angle where total internal reflection occurs
- refraction index of glass (n=1.5) > water (n = 1.3)
- total internal reflection → light does not enter the sample (majority of fluorescent molecules aren't illuminated)




- electromagnetic energy does extend (=evanescent field) for very short distance (100 -200 nm) beyond the surface → this allows the **excitement of just those molecules in the layer closest to the surface** → their emitted light is no longer competing with out-of-focus light from the other molecules and can now be detected

### TIRF IN A PALM EXPERIMENT

- minimize both **autofluorescence and detector noise**, images are taken in a TIRF-microscopy setup!
- The less noise from other sources is detected by the camera, the more reliably it can pick up the photons emitted by the molecules of interest
- Without using a TIRF microscope, the positions of the individual fluorophores could not be determined at high-enough accuracy such that the **advantages of the PALM method would be undone**

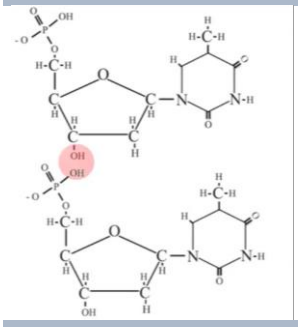
# DNA SEQUENCING

## DNA SEQUENCING WITH THE SANGER METHOD

### DNA AS STORAGE MEDIUM

- **base sequence has almost no influence on the physico-chemical properties of a DNA molecule** or on the speed, efficiency and accuracy of its prosses by enzymes
- 2 consequences of **uncoupling of Information content from the physical properties** of the storage medium
  - For every DNA the same DNA-sequencing methods can be used
  - base sequence of DNA as primary interest → most relevant biological information about a piece of DNA (physical properties and structure of this piece of DNA convey little additional Information (Histonfaltung)

### DEHYDRATION SYNTHESIS



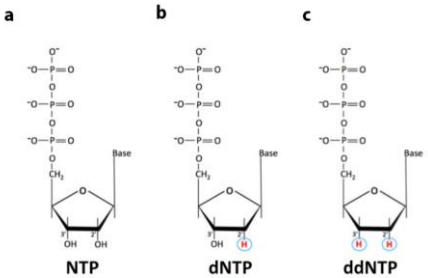
### SANGER SEQUENCING

- was the first reliable and universal method for determining the base sequence of a strand of DNA
- named "dideoxynucleotide chain-termination method"

- needed components for DNA synthesis reaction
  - single-stranded template DNA
  - a primer
  - a polymerase
  - triphosphate deoxynucleotides dATP, dGTP, dTTP, dCTP
- additional for sanger method
  - **dideoxynucleotides (ddNTP)**

#### NUCLEOTIDES

a) building block for RNA
b) building blocks for DNA (H-Atom instead of OH-group increases chemical stability against hydrolysis)
c) synthetically generated **chain terminators** (3'-hydoxyl group of sugar ring is replaced by H-Atom → new nucleotide cannot attach the DNA-strand anymore)

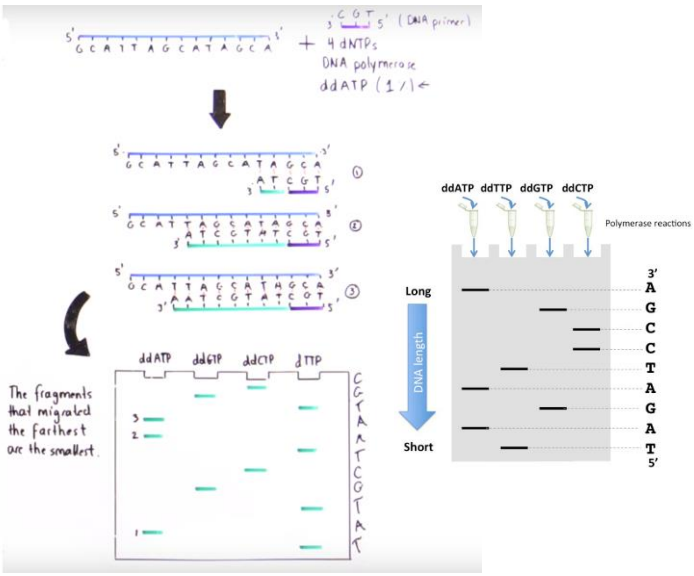



- **α-phosphate group** of the new nucleotide **attaches to 3'-hydroxyl group** of the previously incorporated nucleotide but the lack of 3'-hydroxyl group of ddNTPs prevents the addition of any further nucleotides to the growing DNA strand → **Termination**

#### PROCESS:

- for each probe we just use one kind of ddNTP
- polymerase can either
  - incorporate a natural dATP and synthesis will continue with the next nucleotide
  - incorporate an artificial ddATP in which case the synthesis of this particular strand is terminated
- Which of the two happens is **determined by chance and by the relative concentration** of dATP and ddATP
- The result of this synthesis reaction is a **population of DNA strands of different lengths**

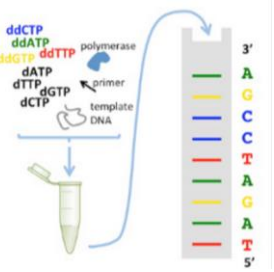
- second step is to read out the information by **separating the newly synthesized strands by gel electrophoresis**
  - shorter DNA strands will move through the gel more quickly than longer strands
  - visualisation of DNA bands on the gel by labeling ddNTP molecules radioactive (later by fluorescence)




- by stepping along the staircase from the bottom of the gel (shortest strand) to the top, the sequence of the newly synthesized strand can be read off in the 5'-to-3' direction →sequence that is read out from the gel is complementary to the originally sequence DNA strand

NEW:

- radioactivity-based version was **replaced by a fluorescence-based version**
  - each of the four dideoxynucleotides is labeled with a different fluorophore
  - makes it possible to combine the four previously separate synthesis reactions into a single reaction and the separation on a single lane of a gel
  - color-sensitive fluorescence detector allows **automatic reading out** the base sequence and automatically determination of the DNA sequence is interpreted by computerized "base-calling" algorithms



## LIMITATION AND ERRORS:

- 1970's
  - labor intensive
  - involved radioactive compounds
  - read length was less than 100 bases
  - determine a few hundred bases per day
- fluorescence, automation of all aspects of the technique
  - efficiency was improved
  - sequence 1mio bases/ day (2'000 - 3'000 US dollars)
  - read length of **1000 base** (in its most optimized form)
- maximal read length is fundamentally limited by the **relative size resolution obtainable by gel electrophoresis** → separation of two strands that are 999 and 1000 bases long, brings the required relative resolution to 1/1000, which is the current technical limit of gel electrophoresis)
- error rate: one error per 10'000 - 100'000 base pairs → **extremely low**

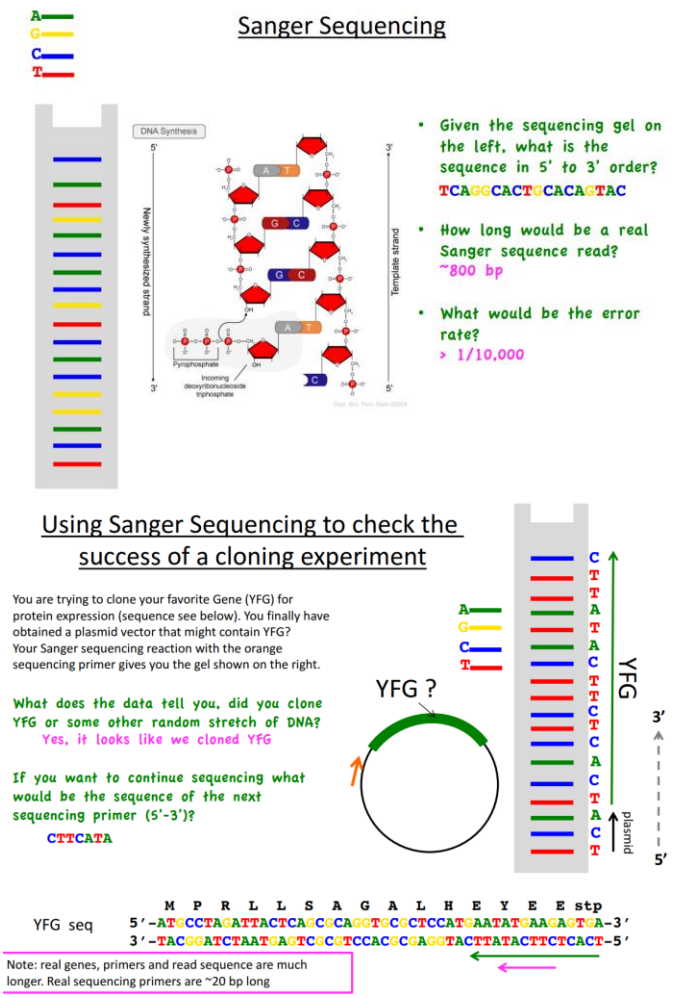
## APPLICATION (=GEBRAUCH/ANWENDUNG):

- Sanger sequencing remains popular for many day-to-day applications in molecular-biology laboratories and for medical diagnostics

PROS:

- **primers that can be chosen** by the experimenter →allows **targeted sequencing** of a specific portion of a larger DNA molecule (2nd and 3rd generation sequence entire DNA)
- **can read longer stretches of DNA** in a single reaction with **very low error rates**

## ERGÄNZUNGEN

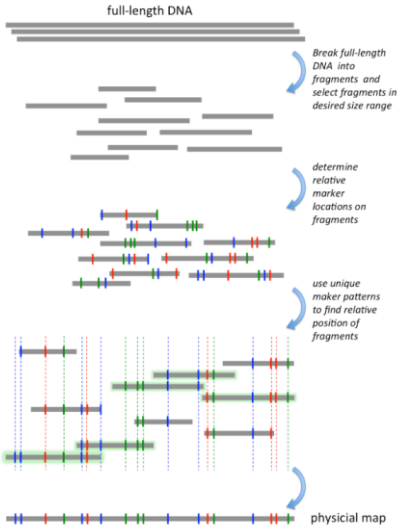


## SEQUENCE ASSEMBLY

- Sanger method can analyse sequences of up to 1000bp
- If longer pieces; **sequencing strategy:**
  - **library generation**: full-length stretch of DNA must **be broken up into parts** that are of the appropriate size for sequencing
  - **sequence assembly**: after sequencing each of smaller pieces, put sequences back together in the correct order to obtain the sequence of the full-length DNA
- two prototypical strategies for sequencing long sequences (whole genomes)
  1. **map-based sequencing**
  2. **shotgun sequencing**

## 1. MAP-BASED SEQUENCING (BAC-TO-BAC)

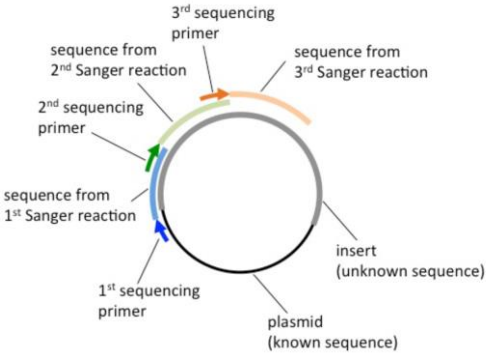
- initial, very long piece of DNA is systematically broken up into consecutively smaller fragments while mapping the exact location of each of these fragments
- result: library of DNA fragments where **each fragment is stored in its own test tube** and the **location of every fragment** in the initial full-length piece of DNA **is known**
- fragments are then sequenced individually
- resulting fragment sequences can be assembled in their known order to obtain the sequence of full-length DNA




1. break the full-length DNA into large fragments of about 100'000 to 200'000 bp length by **restriction enzymes or physical shearing**
2. large fragments are inserted into **bacterial artificial chromosomes** (**BAC**s= circular DNA constructs, which allow the insertion and propagation of long stretches of DNA in a similar way that bacterial plasmids allow the propagation of smaller DNA fragments)
3. position of each large fragments in the initial full-length DNA sequence is determined by **physical mapping** (physical map indicates the position of sequence-specific markers)
4. Selection of subset of BACs that covers entire sequence
5. Break up the selected BACs and isolate & process the insert further until fragments are about 10'000 bases long
6. shorter fragments are cloned into plasmid vectors
7. Again, physical mapping (used to determine the position of each smaller fragments within their respective BAC)
8. again, a subset of fragments spanning the entire BAC is selected for further analysis
9. 10'000-bp fragments in the plasmids are small enough to be sequenced using a **primer-walking** approach

### PRIMER WALKING:

- can be used to sequence stretches of DNA that are several times longer than the maximal read length of Sanger sequencing (10'000-bp)




- initial sequencing reaction: uses a **primer** that is complementary to a known portion of the plasmid (5' end)
- after determination of the first unknown sequence, the now known **end** can be used to generate **the primer** for the next sequencing reaction etc.

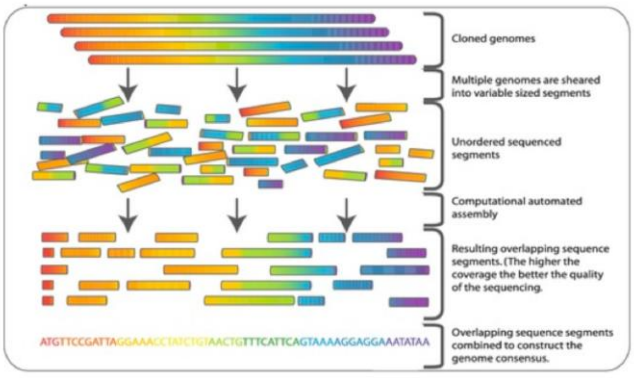
PROS/CONTRAS

- labor-intensive
- expensive
- logistically challenging
- very straightforward process → individual fragment sequences can be assembled unambiguously (eindeutig, clearly) into the full-length genome sequence
- repetitive Sequenzen stellen bei der BAC-to-BAC Methode ein kleineres Problem dar → primer walking

## 2. SHOTGUN SEQUENCING

- no subcloning, mapping, and primer-walking steps

1. full-length piece of DNA is cut by sonication (=Ultraschallbehandlung) into fragments that are short enough to be sequenced in one continuous read
2. fragments are then sequenced individually
3. Multiple copies of fragments **will overlap partially** (have identical sequences as their neighbours) → can be used to assemble the full-length sequence
   overlap of just 16 bases ($4^{16}$ = 4.3 billion combinations) would be sufficient to uniquely identify a neighbour




- simplifies the library preparation and sequencing process
- but the resulting **library is completely unordered**

### PROBLEM OF REPEAT SEQUENCES

- many DNA sequences of interest (whole mammalian genomes) contain **duplicated genes** and **long stretches of repeat sequences**

- repeated sequences generate **ambiguities** (=Mehrdeutigkeiten/Unklarheiten) in the assembly of shotgun sequencing data
- it is typically not possible to assemble entire mammalian genome using a pure shotgun approach → needs to be complemented by **traditional mapping** data or **special paired-end** or **mate-pair sequencing data**

## RE-SEQUENCING VS. DE NOVO SEQUENCING

- ***de novo* sequencing**: sequencing DNA with a completely unknown sequence
  - map-based sequencing
  - shotgun sequencing (or combinations of the two)
-
- ***"re*-sequencing"**: using a **known, highly similar sequence as a template** that guides the assembly of a short sequence → genome sequenceof a different individual from the same species is **already available**
  - today the most sequencing projects
  - single-nucleotide polymorphisms, short insertion or deletion mutations are **relatively easy to detect** and map → overall structure of the template and the new sequence are essentially identical
  - large-scale structural rearrangements and elongation or shortening of regions with repeat sequences are hard to detect → template sequence is not correct

### ANWENDUNGSBEISPIEL

**Huntington's disease**

- Autosomal dominant inherited disease
- Early onset neurodegenerative disease
- Incidence ~5 in 100'000
- No cure

brain with Huntington's disease

healthy brain

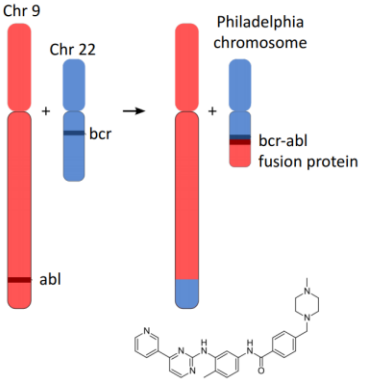
Polymorphisms in the 1st exon of the HTT gene

```
ATGGCGACCCTGGAAAAGCTGATGAAGGCCTTCGAGTCCCT
CAAGTCCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC
AGCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCCGCCACCG
CCGCCGCCGCCGCCGCCGCCTCCTCAGCTTCCTCAGCCGCC
GCCGCAGGCACAGCCGCTGCTGCCTCAGCCGCAGC
```

CAG=Gln
poly glutamine tract

| # repeats | classification | prognosis | risk to offspring |
|---|---|---|---|
| <26 | normal | will not be affected | no risk |
| 27-35 | intermediate | will not be affected | increased risk (<<50%) |
| 36-39 | reduced penetrance | may be affected | 50% |
| 40+ | full penetrance | will be affected | 50% |

Detecting large chromosomal rearrangements




**Chromosome translocation creates the new fusion protein *bcr-abl* – a permanently activated tyrosine kinase that causes cancer (chronic myelogenous leukemia).**

**How would we see such a rearrangement in 2nd gen sequencing data?**
Reads that bridge the chromosome to chromosome fusion sites.

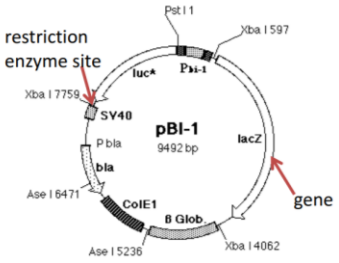
**Why might we miss it?**
Assembly software may reject reads as "contamination" because of poor match to template sequence

**What type of sequencing data, or other data could we use to detect this rearrangement?**
Paired end or Long-read sequencing data, mapping data (e.g. BioNano)

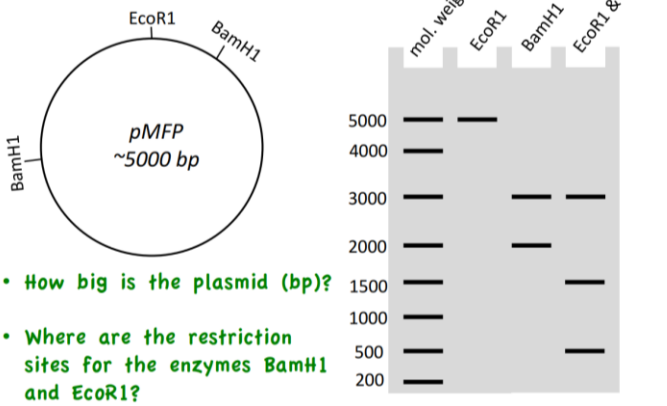
# ERGÄNZUNGEN DER VORLESUNG

## MAPPING

- provide an overview
- organize and integrate knowledge
- identify unknown areas
- useful even when not perfect
- a hypothesis that can be improved over time

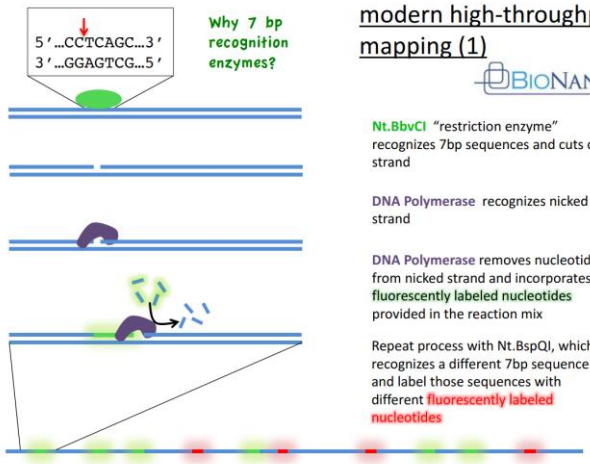



Very simple example for mapping:
Restriction-digest-based mapping




- How big is the plasmid (bp)?
- Where are the restriction sites for the enzymes BamH1 and EcoR1?

- We use restriction maps of fragments in plasmids to generate map of original DNA sequence!
- Generation of restriction maps of plasmids: Multiple copies of an original linear DNA sequence were broken into fragments and then cloned into plasmids
- *Can you generate a map of the original DNA sequence*? YES




modern high-throughput mapping (1)

BIONANO

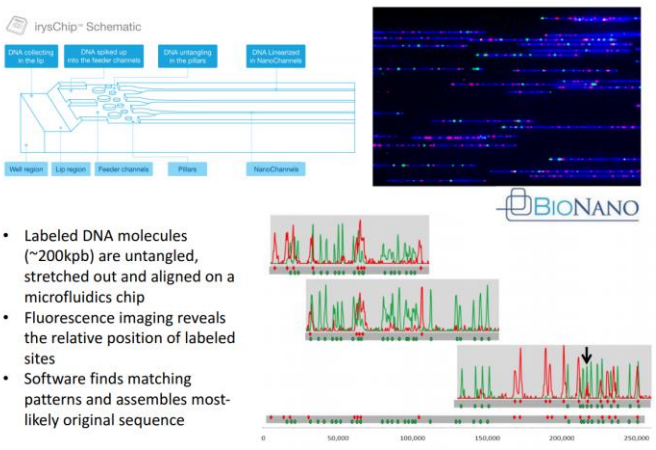
Nt.BbvCI "restriction enzyme" recognizes 7bp sequences and cuts one strand

DNA Polymerase recognizes nicked strand

DNA Polymerase removes nucleotides from nicked strand and incorporates fluorescently labeled nucleotides provided in the reaction mix

Repeat process with Nt.BspQI, which recognizes a different 7bp sequence and label those sequences with different fluorescently labeled nucleotides

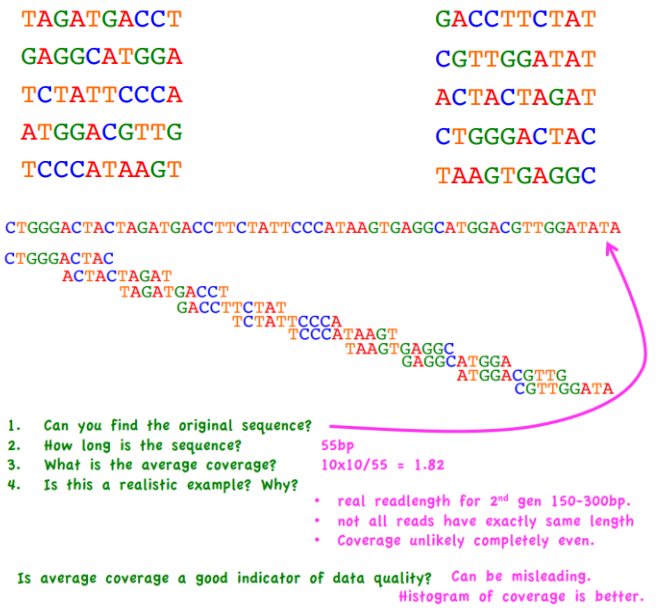
modern high-throughput mapping (2)




- Labeled DNA molecules (~200kpb) are untangled, stretched out and aligned on a microfluidics chip
- Fluorescence imaging reveals the relative position of labeled sites
- Software finds matching patterns and assembles most-likely original sequence

# ASSEMBLY OF DANN

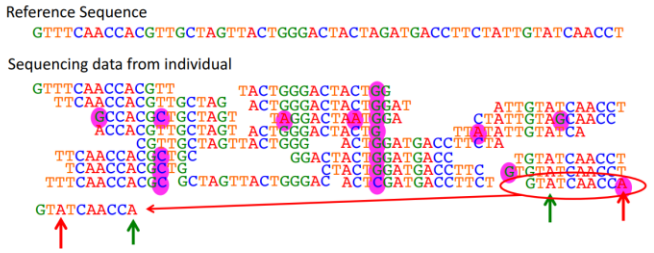
## SHOTGUN SEQUENCING DATA:




1. Can you find the original sequence?
2. How long is the sequence? 55bp
3. What is the average coverage? 10x10/55 = 1.82
4. Is this a realistic example? Why?
   - real readlength for 2nd gen 150-300bp.
   - not all reads have exactly same length
   - Coverage unlikely completely even.

- Is average coverage a good indicator of data quality? Can be misleading. Histogram of coverage is better.

Assembly against a reference sequence: Distinguishing **S**ingle **N**ucleotide **P**olymorphisms (SNPs) from sequencing errors




- How many nucleotides in the aligned sequence reads differ from the reference? 19
- What type of deviations do we see? Single nucleotide substitutions
- What other deviations would be possible? Insertions, deletions
- Are those deviations sequencing errors or what other explanation could there be? SNPs
- Can you imagine a situation where sequencing errors and alignment choice interact?
- How could you estimate the sequencing error rate from the data you have? Unaccounted deviations / nucleotides sequenced ≈ 2%

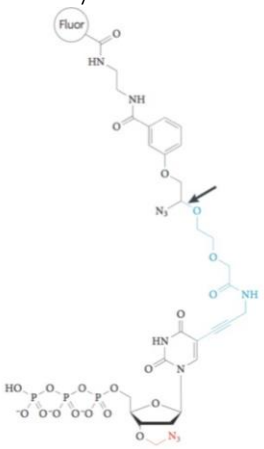
Here are the short-read data of two patients with neurological problems assembled to the reference sequence of the HTT gene. Do you find evidence for a possible mutation? Maybe repeat stretch in patient B is much longer than in reference sequence

## SECOND-GENERATION SEQUENCING

- early 2000's: rapid succession of sequencing technologies → group of technologies = "next-generation sequencing"
- First-generation technologies: (e.g. Sanger etc.)
  - clear **separation** of the **sequencing reaction** and the **read-out process**
- Second-generation technologies:
  - **No separation** → one integrated process that takes place in the same reaction vessel
- need for an **amplification step** → PCR-like process prior to the actual sequencing reaction
- Third-generation technologies
  - **no amplification step** → manage to sequence individual molecules from the sequencing libraries directly
  - much longer sequence reads are possible

## SEQUENCING-BY-SYNTHESIS (SBS) METHOD

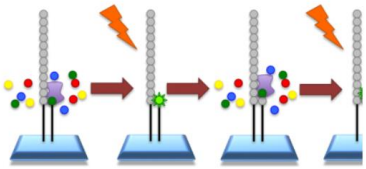
- dominant second-generation sequencing technology
- increased the speed and reduced the cost of sequencing by factor 10'000
- similarity to Sanger method: uses a primer, a polymerase, a template strand, and nucleotides for the synthesis of a new DNA strand
- main difference: type of used nucleotides
  - Sanger: mix of natural and modified nucleotides and functional groups are attached permanently
  - SBS: **only uses modified nucleotides** and the functional groups of the nucleotides (fluorophore) can be chemically cleaved




Fluorescently labelled, chain-termination dCTP
- Azide group prevents incorporation of additional nucleotide and can be chemically removed to generate a 3'hydroxyl group
- Chemically cleavable linker attaches the fluorophore (each type of base carries a different colour)

### PROCESS:

1. Isolation of DNA from a specimen (5ug DNA)
2. DNA is mechanically sheared into 500-1000bp fragments
3. DNA fragments are ligated **to adapter oligos containing ID-tag, surface-attachment and primer-complementary regions**
4. Library molecules are **bound to solid surface of flow-cell** so that individual molecules are physically well separated
5. Individual **bridge amplification** to generate clusters of identical molecules
6. flow cell is flooded with solution **containing the four different types of fluorescently labeled,** chain-terminating nucleotides
7. polymerase uses the template strand as a guide to extend the primer with the complementary nucleotide → The terminating group on the newly incorporated nucleotide prevents the incorporation of additional nucleotides
4. All non-incorporated nucleotides are washed away
5. Measure of the **fluorescence signal** shows, which of the four nucleotides has been built-in
6. Both the **fluorophore and the terminating group are removed** by chemical cleavage and washed away → prepares the reaction for the next cycle

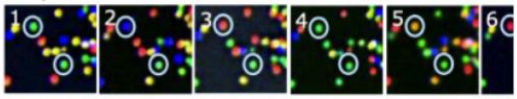


Optional steps for paired-end sequencing:
- The newly synthesized strand is released and the opposite end of the template is attached to the surface
- Sequencing-by-synthesis is repeated to obtain the sequence of the opposite end of the fragments

### PROGRESS:

tremendous performance gains (Leistungssteigerung) through **parallelization** and **miniaturization**

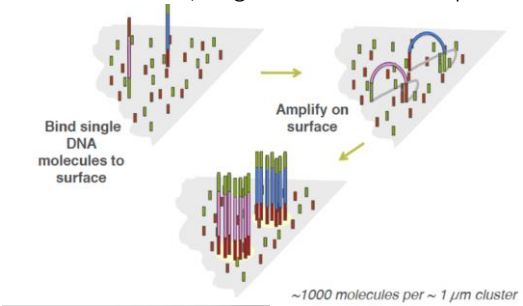
- **no improvement of speed or accuracy** of the individual reaction but **SBS can be parallelized and miniaturized**
  - billions of parallel sequencing reactions in a flow cell the size of a microscope slide
  - key change enabling this performance is that the DNA fragments are **attached to a solid surface** → individual reactions can be identified by their **x-y-coordinates** → this makes it possible to perform many individual sequencing reactions side-by-side
- biochemical reaction and physical read-out are one continuous process without physical transfer




### BRIDGE AMPLIFICATION:

is used to generate clusters of identical molecules to yield a detectable optical signal

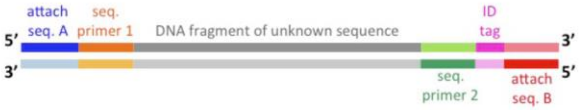
- every visible coloured dot on the flow cell surface corresponds **to a cluster of DNA molecules** (not to a single molecule!)
- signal from a single fluorophore molecule would be too weak to be detected reliably → clustering!
- bridge amplification (a PCR-like process) is used to amplify individual surface-bound DNA molecules into clusters of sequence-identical molecules
  - occurs before the sequencing-by-synthesis reaction
  - end of the DNA molecules always remains attached to the flow-cell surface
  - A few cycles generate a tight cluster of approximately 1000 identical, single-stranded DNA copies



## ADAPTER SEQUENCES:

Enable surface attachment and ID-tagging of DNA-fragments

- Problem: DNA pieces have different sequences and all sequences are unknown → synthetic adapter oligos are enzymatically attached to the ends of the unknown DNA sequences → are identical for each of the DNA → same sequencing primers and surface-attachment oligos can be used for each DNA molecule
- Adaptor sequences (red and green) are attached to the end of initial DNA and bind complementary primers that are attached to the flow cell's surface
- Primer are extended using a DNA-Polymerase reaction
- Heat denaturation separates the newly formed double stranded DNA molecule
- Process is based on **sequence complementarity between the oligos and primers**
- ID facilitates the Identification of the sample, from which the unknown DNA fragment derived




## PERFORMANCE CHARACTERISTICS

- Several billions of individual DNA fragments can be sequenced in parallel with read lengths of up to 300 bases and an error rate of approximately 0.1%

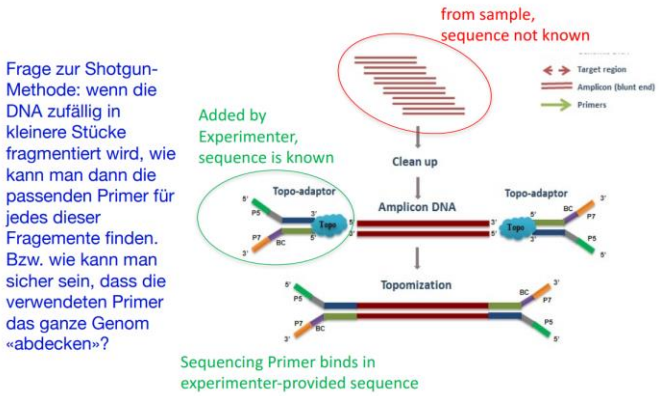
## LIMITATION:

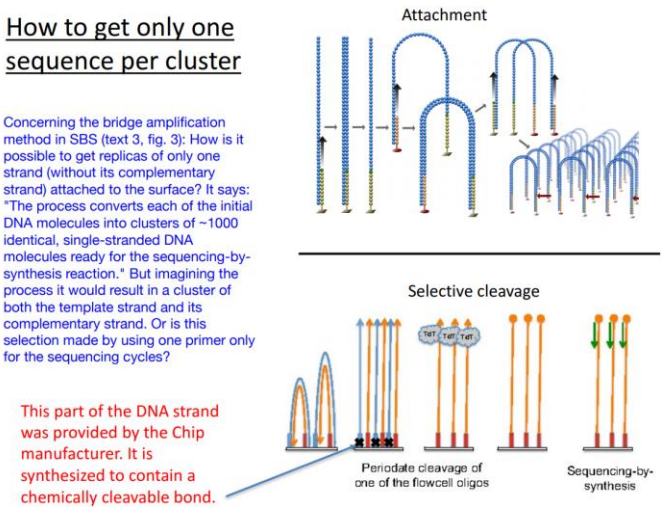
need to synchronize limits the **speed and read-length**

- **very slow!** → ~1 hour to add one additional base (vs. DNA polymerase: ~1000-bases-per-second)
- slowness due to the need of **perfect synchrony** between the reactions of the several thousand identical molecules in each of the clusters (Without synchrony, fluorescent signals of different molecules in a cluster would differ)
  - **substantial waiting** time at each reaction step to ensure all molecules have completed the reaction
  - **time-consuming flushing-in and washing-out**
- synchrony also limits the maximal read-length → the longer, the more molecules are out-of-lockstep

## ERGÄNZUNGEN

### LIBRARY PREPARATION:



## THIRD-GENER ATION SEQUENCING

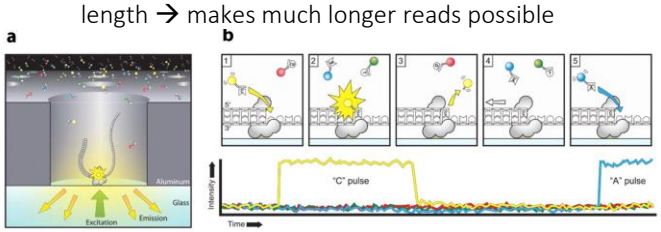
sequencing individual molecules

- solves the synchrony-based problems by performing the sequencing reaction on individual molecules

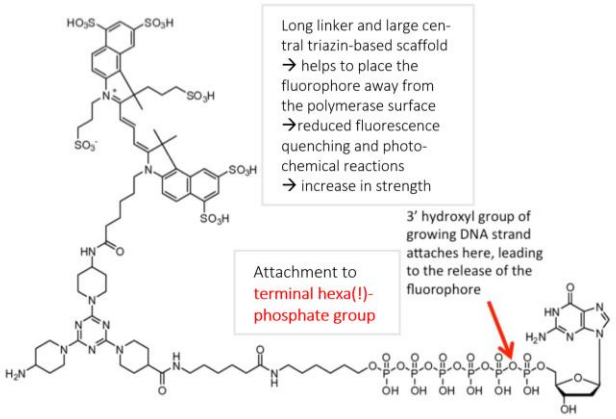
## PACIFIC BIOSCIENCES (PACBIO)

Third-generation fluorescence-based single-molecule real-time (SMRT) sequencing

- performing a **sequencing-by-synthesis-style reaction on individual DNA molecules**
- all necessary reagents are added and nucleotides are observed in real-time while being inserted
  - simplifies the experimental protocol
  - potential to greatly speed up
  - removes the synchronicity-imposed limits on read length → makes much longer reads possible




Individual fluorescently labelled nucleotides → during the time between successful binding and the formation of the photodiester bond, the gamma phosphate linked fluorophore is excited and emits an optical signal that is picked up by the instrument's zero-mode wavelength (ZMW) optics



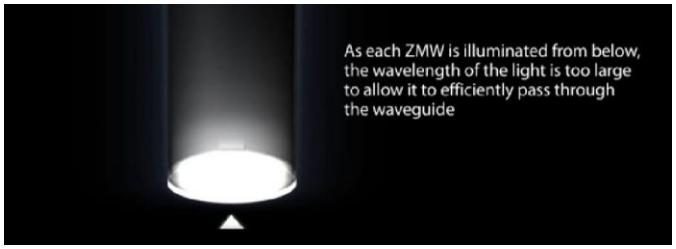

### OPTICAL MONITORING:

**extremely small reaction volumes** makes single-molecule sequencing possible

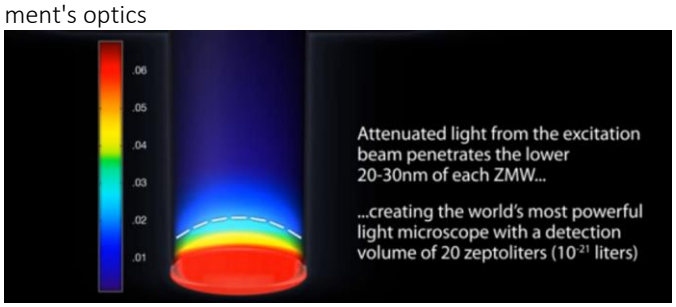
- signal obtained from **a single fluorophore** is **limited**!
  →Boost the signal
- fluorescence background
  →minimizing the fluorescence background from the other fluorescently labelled nucleotides diffusing in the solution
- method: **monitor fluorescence only in extremely small volume** of a few tenths of zeptoliters (1zeptoliter = $10^{-21}$)
- ZMV provides 1000-fold improvement in the reduction of background noise

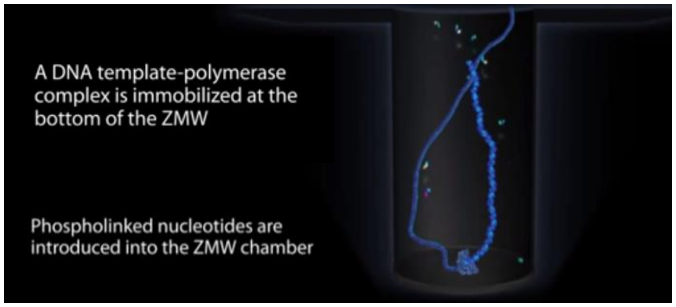
### SMRT TECHNOLOGY AND ZMW:

- SMRT cell enables the potential detection of thousands of sequencing reaction in parallel
- Each SMRT cell contains tens of thousands of zero-mode waveguides (=ZMW)
- ZMW provides the world's **smallest light detection volume**
  → special optical configuration caused by combination of size (much smaller than the wavelength of light) and special chip material

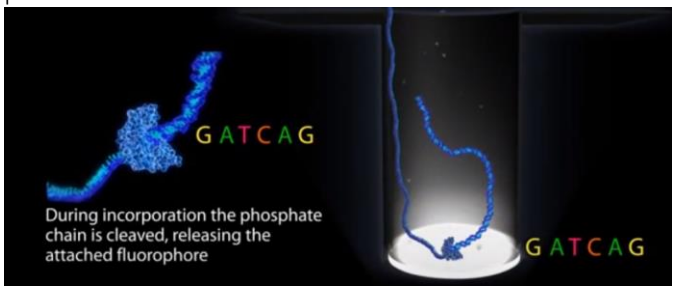



- Light impinging on a ZMW is **not able to traverse the chip**, so no **light leaks into solution** above the well but light generates an **electromagnetic field** which can excite fluorescent molecules that are located less then ~50nm from the bottom of the well
- photons emitted by those fluorescent molecules can also "escape" through this ZMW to be detected by the instrument's optics




- Each of the four nucleotides is labelled with a different coloured fluorophore
- As a base is held in the detection volume, a light pulse is produced

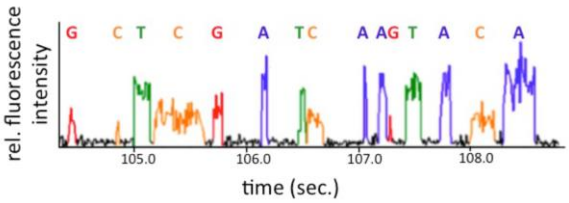



### SPEED PACBIO VS. SEQUENCING-BY-SYNTHESIS:

- **Faster reaction speed**
  - adds several nucleotides per second
  - ~10'000-fold improvement in speed
- **lower degree of parallelization** limits the throughput
  - PacBio allows parallelization of "only" a few thousand compared to billions of parallel reactions by S-B-S
- PacBio technology **still lags behind** the sequencing-by-synthesis technology in terms of **overall throughput**

LIMITATION AND IMPROVEMENT:

- **increased error rates** → lies around 10%
  - weak signals from single-molecule detection
  - difficult interpretation of time-traces (large variation in the duration from peak to peak)




- **no advantage in the absolute throughput**
- advantage in **achievable read length** →can generate continuous reads of >60'000 bases
  - no synchronization problems that limit the read length to ~300 bases
  - read-length is limited by very rare (1 per ~15'000 bases) **photochemical reactions, in which a photoexcited fluorophore reacts with polymerase and destroys it**
- (occurs randomly, often around ~10'000 bases, but more than 60'000 bases possible → very long reads are valuable for the assembly of difficult regions of a sequenced genome as repeats or gene duplications
- Downside: **inherently noisy**

USAGAE:

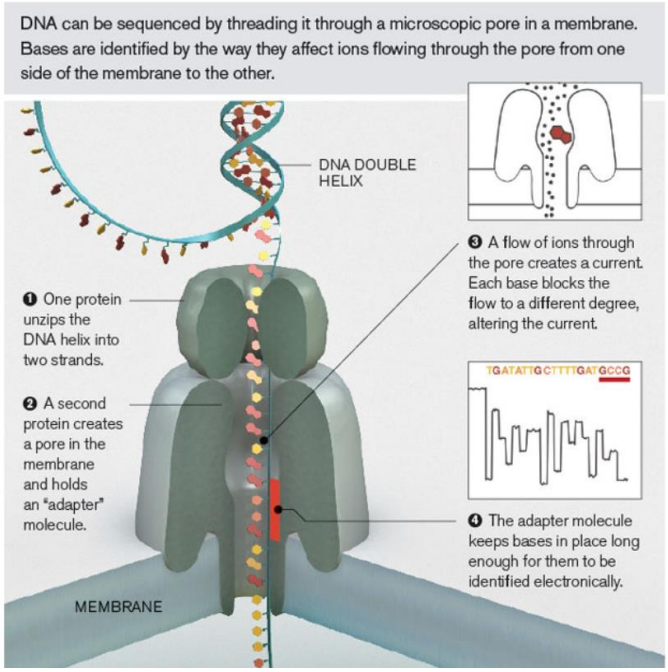
- Cause to **high error rate** its main usage lies in providing a **long and continuous sequence template** onto which shorter, lower-error reads from other techniques such as sequencing by synthesis can be assembled

## NANOPORE SEQUENCING

- sequencing methods before are all based on:
  - **DNA polymerases** (enzymatic rate limitation)
  - **fluorescently labeled nucleotides** (limitation: spatial resolution, number of excitation-emission cycles by photochemical damage, very high cost)
- Nanopore sequencing requires neither DNA synthesis nor fluorescently labeled nucleotides

MECHANISM:

- **reads sequence of a DNA strand directly**
- protein nanopore is embedded in a membrane that separates two chambers containing electrolyte solutions. A **voltage** applied across this membrane pulls ions and the DNA strand through the pore. The presence of the DNA in the pore restricts the **flow of electrolyte ions** in a way that depends on the base sequence of the DNA section that is currently traversing the pore. By **measuring and analysing the electrolyte current** over time, the sequence of the DNA passing through the pore can be determined




- Due to slightly different size, shape and polarity of the different bases, they have slightly different effect on electrolyte current
- pore seems to be a genetically modified version of a **porin protein** (span outer membrane of many bacteria)

CHALLENGE:

- current depends not only on the base itself, but also on the **identity of the neighboring bases**

LIMITATIONS AND ADVANTAGES:

fast, ultra-long sequence reads in a small portable instrument

- has **no read-length limit** (up to 200'000 bp)
- -) error rate appears to be ~ 10% (or slightly higher)
- **+) speed** ~500 bases per second (close to DNA-synthesis-based sequencing technologies)
- **+) precise, cheap and miniaturized**
  - cheap-to-manufacture and small semi-conductor chip
  - no need for fluorescence-based instruments and expensive fluorophore nucleotides

## SBS VS PAC BIO

**SBS**

Errors in the reaction of individual molecules are not an immediate problem because there are many molecules. But, the effect of these error accumulates and leads to continual and irreversible degradation of the signal.

To be or not to be that is the question
To be or not to be that is the q

**Pac Bio**

Because the signal comes from a single molecule, every error in a reaction immediately leads to an error in the sequence. But a single molecule cannot "get out of phase" with itself. Therefore the signal does not degrade over time.

To be or not to be that is the question
To be ar not to b tht is the quesstion

# PROTEIN MASS-SPECTROMETRY

## MASS SPECTROMETRY (MS)

- very broad range of techniques and instrumentes
- **peptide-fragmentation based analysis of protein samples** as a powerful and widely used experimental technique

## PRINCIPLE & COMPONENTS OF MASS SPECTROMETER

mass spectrometry determines the mass of individual molecules by **measuring how electrically charged versions of these molecules (ions)** move when exposed to electric or magnetic fields in a vacuum

COMPONENTS:

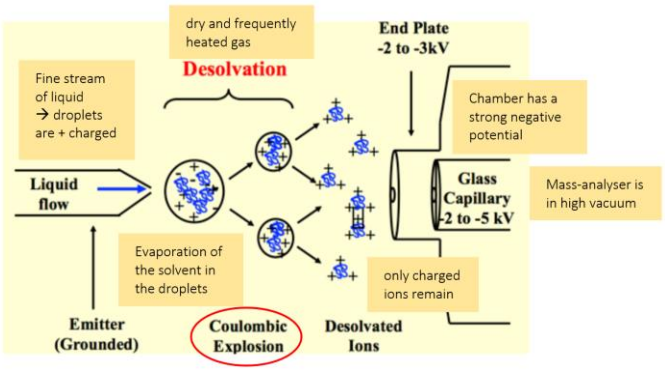
- **ion source** → deposits an electrical charge on the molecule and transfers it from solution into the gas phase
- **mass analyzer** → separates different ions based on their mass-to-charge ratio (*m/z*)
- **detector** → counts the ions passing the mass analyzer

ION SOURCES:

- Biological molecules usually **exist in aqueous solution**
- For analysis in mass spectrometer, molecules need to be
  - **charged**
  - **isolated** from the surrounding solvent molecules
  - transferred into the **gas phase**

→ ion source accomplishes these three steps

- two main types of ion sources
  - matrix-assisted laser desorption ionization (MALDI)
  - **Electrospray ionization (ESI)** → for analysis of peptides from protein samples




1. ESI ion sources spray very fine stream of liquid out of a nozzle → droplets carry an excess **positive charge**
2. on the way to mass analyser, the droplets pass through a zone of **dry gas** → solvent in these **droplets evaporates**
3. this concentrates the charges on the droplet surface causing **"coulombic explosions"** that shatters large droplets into smaller droplets, which again shrink due to evaporation and explode again
4. process continues until the solvent has evaporated completely and only the charged ions of the sample molecules remain

**MERKE**: only charged molecules continue their flight path while solvents, neutral fragments and other gasses are removed by vacuum pumps

MASS FILTERS/ANALYZERS:

- filtrates passing ions
- lets accelerated ions only pass, if **these ions have a particular m/z ratio** that was specified by the experimenter

DETECTORS:

- **records the arriving ions**
- small number of arriving ions → signal needs to be amplified by an electron multiplier before it can be measured

## MASS SPECTRUM

= plot of the **relative intensity of the detector signal** as a function of the ***m/z* ratio of the ions** generating that signal

shows the relative abundance of ions with different *m/z* ratios

- **one peak** does **not** correspond to **one molecular entity**!
  - Multiple chemical entities may contribute to one peak
  - one molecular will typically generate multiple peaks

## ISOBARIC SPECIES

**isobaric species =** molecular species with exactly same mass

- same atomic components
- different arrangement, structure and significance
- if same atomic composition (and charge) they contribute to the **same peak** in the mass spectrum regardless of how the atoms in the molecule are arranged

## MULTIPLY-CHARGED IONS

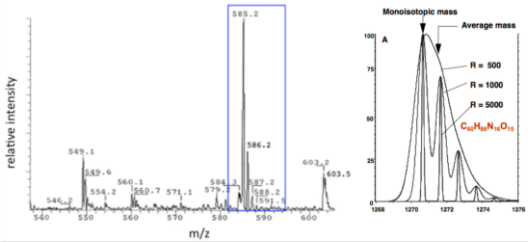
- A single type of molecular can **generate multiple peaks by adopting multiple charge states**
  - received different charges in the **ionization phase**
  - mass spectrometer measures **mass-to-charge ratio** → doubly charged ion will appear at roughly half the *m/z* ratio of the singly charged ion
  - actual *m/z* for doubly charged ion = **$[M+2H]^{++}/2$**

## ISOTOPES

- Isotopes = atoms with the same number of protons and electrons but **differ in their number of neutrons**
  - same chemical properties
  - have different masses

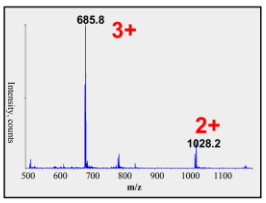
| Atom | Mass | Rel. abund. (%) |
|---|---|---|
| **carbon** | 12.000 | 98.900 |
| | 13.003 | 1.100 |
| **hydrogen** | 1.008 | 99.985 |
| | 2.001 | 0.015 |
| **nitrogen** | 14.003 | 99.630 |
| | 15.000 | 0.37 |
| **oxygen** | 15.995 | 99.760 |
| | 17.999 | 0.200 |
| **sulfur** | 31.972 | 95.020 |
| | 33.968 | 4.210 |

- isotopes of carbon are important!
  - carbon is present in virtually all biomolecules
  - possesses a relatively common isotope (1/100)
  - $^{13}C$ atoms have the same chemical properties as $^{12}C$ –
  - group of chemically identical molecules will contain some versions of $^{13}C$ atoms → have a higher mass
- **isotope ladder** visible in a mass spectrum
  - chemically pure compounds will consist of a mix of molecules with different masses
  - the number of $^{13}C$ atoms in a molecule follows **a Poisson distribution**
  - Serie of mass peak increases by one mass unit and decreases in intensity

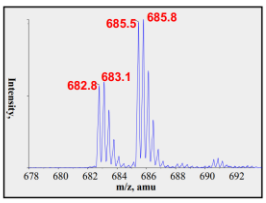



- Due to these isotope ladders, a mass spectrometer needs to have **a sufficient resolution**
  - Low resolution: peaks will **fuse to one big peak** →the resulting aggregate peak will have a mass value that is **hard to interpret**

Multiply Charged Ions

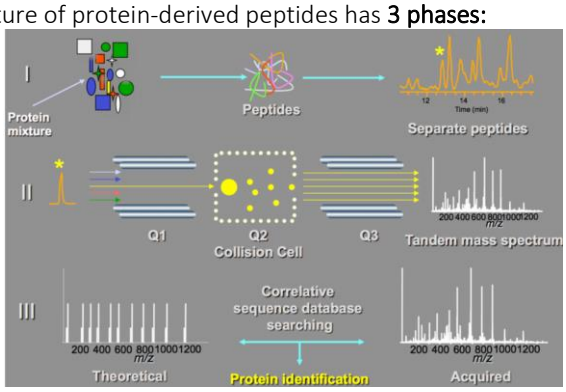



Isotope Distribution



## WORKFLOW

workflow for using **MS/MS to identify peptides** in a complex mixture of protein-derived peptides has **3 phases:**




I) **sample preparation** (extraction and digestion) and **pre-fraction** of peptides by HPLC

II) **MS/MS in a mass spectrometer** equipped with 2 mass filters (Q1 and Q3) and an interposed collision-based fragmentation stage (Q2)
(Peptide sequences are identified by MS/MS)

III) **Identification of individual peptides** by matching their observed fragmentation ion spectrum to a database of all theoretically possible fragmentation spectra, predicted based on the sequence of the organism's protein-coding genes

## I.1) SAMPLE PREPARATION

- proteins need to be extracted from the tissue and prepared for analysis in the mass spectrometer

1. extracting proteins from samples by disruption of the tissue by physical means (blender, cryo-pulverization…)
2. chemical extraction steps using detergents and salts
3. proteins in the samples are denatured
4. disulfide bonds are chemically reduced (prevent reforming disulfide bonds by blocking cysteine residues)

## I.2) PROTEIN DIGESTION

- it's technically possible to analyse the masses of intact proteins in a mass spectrometer (but it's rarely done)
- **proteins are digested into peptides prior MS analysis**

1. **digest proteins into peptides with a length of ~10-20aa**
2. analyse these peptide mixtures by mass spectroscopy
3. construct the composition of the initial protein sample based on the observed peptide mix

PROTEASE: well-controlled enzymatic digestion step

- protein digestion is performed by **specific proteases** →
- most common protease is **trypsin**
  - cuts the peptide backbone of a protein on **C-terminal** side of an **arginine** or **lysine** residue
  - leads peptides with size of around 10-20 amino acids
  - sequence specificity provides a useful constraint in the identification of a peptides sequence → C-terminal amino acid has to be either an arginine or a lysine

## I.3) PRE-FRACTIONATE

- pre-fractionation of the digested peptide-mix before injecting it into the mass spectrometer
- achieved through **high-pressure liquid chromatography (HPLC)** of peptide mixture on a reversed-phase column
- outlet of the HPLC column is directly connected to the ESI ionization stage of the mass spectrometer
  - flow rate of HPLC more slowly than MS/MS-analysis → gives mass spectrometer time to analyze mixture of peptides arriving at ionization stage in real time

## II) MS/MS SPECTROMETERS

- possess two or more mass filters in series and between a fragmentation stage

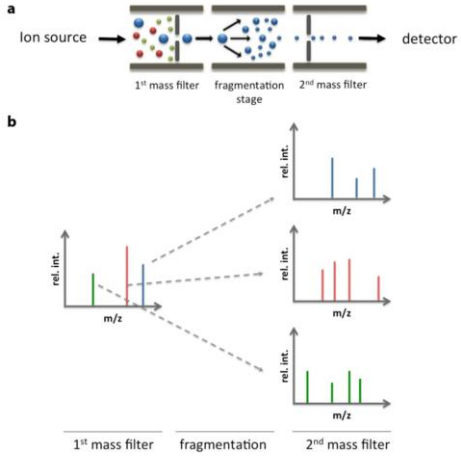
CID:

- **collision-induced dissociation (CID) stage =** popular fragmentation stage
  - fast-moving molecular ions collide with neutral gasses (helium, argon, nitrogen) → kinetic energy of collision is converted into very strong intra-molecular vibrations that can disrupt chemical bonds → molecular ions are broken down into smaller pieces

MS/MS EXPERIMENT

1. first mass filter **selects ions according to its m/z ratio** → ion's m/z value and selection/isolation of one Ion
2. selected ions enter the fragmentation stage and break down into fragments
3. second mass filter analyses smaller fragments → records the mass and intensity of the fragment ions generated from the isolated molecular ion

- fragmentation pattern is often very characteristic



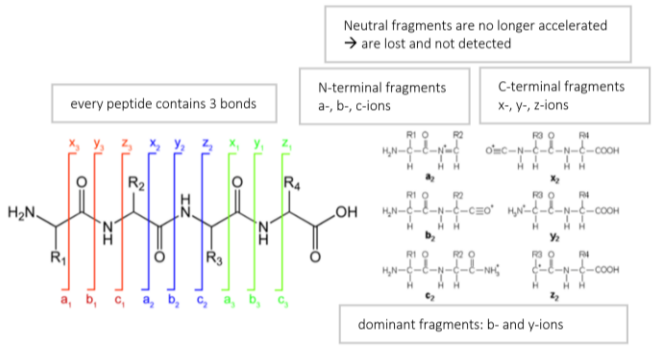

- Problem: number of ion types that arriving at the first mass filter is too large → it is not possible to record fragmentation spectra for every one of the molecular ions

## II.1) FIRST MASS FILTER

- analyses the **composition of the peptide ions** (masses and approximate abundances/amounts) and **selects individual peptide** species for further analysis
  - allows **only one** of these peptides to pass to the subsequent analysis stages (fragmentation)
- a typical tissue protein sample contains too many different types of peptides, so that a typical MS/MS experiment is unable to fragment and analyse all of them
  - usually no major problem, because presence of a protein in a sample can often be detected by the presence of one or two characteristic peptides per protein
  - strategy for selecting depends on the exact scientific question the experimenter wants to answer

## II.2) FRAGEMNTATION

- Fragmentation via collisions of highly accelerated peptide ions with **inert gas ions** inside a collision cell
- conditions are chosen that a peptide ion only experiences **one collision** which results in only **one bond being broken**
- break of peptidebond occurs in the **backbone of peptide** ions (and not in amino acid side chains)




- break occur with comparable, but not identical frequency at each one of the peptide bonds in the peptide → all possible fragments of a peptide can be generated
- mass differences due to individual masses of amino acids
- Fragmentation spectra of peptides are **rarely clean enough to reconstruct the sequence** of original peptide →instead of trying to reconstruct the sequence, one tries to **predict possible peptides**, calculate their fragmentation spectra and **compare** the predicted spectra with the observed fragmentation spectrum
  - need the genomic information of the organism to predict the sequence of the proteins produced by it
  - enzymatic digestion of proteins into peptides is then simulated in the computer, and the masses for the b- and y-ions are calculated

Observed fragmentation spectra can be matched against predicted spectra to identify peptides

- **scoring function**: used to **assess the quality of the match** between the observed and predicted fragmentation → end result is the probability with which the current peptide corresponds to a particular predicted peptide

## ABSOLUTE OR RELATIVE QUANTIFICATION?

- We distinguish between absolute and relative quantification.
  - In the first the absolute concentration of a peptide is determined by adding a calibrated external standard
  - In the second, the relative concentration of one or many peptides is compared in two or more samples

**Absolute quantification**

- Signal (m/z) = *f* (concentration)
- The response factor (F, slope of the calibration curve) is not identical for each analyte, but has to be known for exact determination
- F is dependent on the chemical structure (ionizable groups etc.) – **different peptides have different quantitative response**
- Pure analyte needed as reference for exact determination
- Less accurate large-scale quantitation possible for large data sets

**Relative quantification**

- Measure ***relative intensities / ratios***
- More commonly used to differentiate between biological states

## FROM THE PEPTIDES TO THE PROTEIN

- protein inference = protein identifications from identified peptides
- Identified peptides can be used to conclude the presence of their precursor proteins in the original sample
  - only a small fraction of the observable ions of a protein are actually observed → but number is generally sufficient to infer the presence of a given protein
  - in some cases, a single, highly unique peptide can be sufficient to establish presence of a particular protein
    **= proteotypic peptide**
  - in other cases, several peptides need to be identified to make a reliable statement about the presence of a particular protein
- complicated if:
  - proteins have **high sequence identity** (proteins whose genes are derived from recent gene duplications)
  - if the particular protein has **multiple splice variants**
- process involves a serie of **statistical considerations** that are implemented in specialized **software packages**

## QUANTIFICATION

- measuring the relative or even the absolute abundance (Anwesenheit) of a protein via MS is a **non-trivial task**
- the **intensity** of the detector **signal correlates poorly with the abundance** of the protein in the sample! → intensity of signal is a poor indicator for the peptide's abundance
  - intensity of signal depends on efficiency of different processes (ionization, fragmentation, signal detection)
  - individual efficiencies of processes vary greatly across peptides → overall efficiency can vary across peptides
  - even for the same peptide the efficiency will vary due to **ion suppression**: different peptide molecules compete for a limited number of charges in electrospray droplets → high abundance of a peptide that readily attracts charges may deprive other peptides of a charge & prevent those peptides from being detected

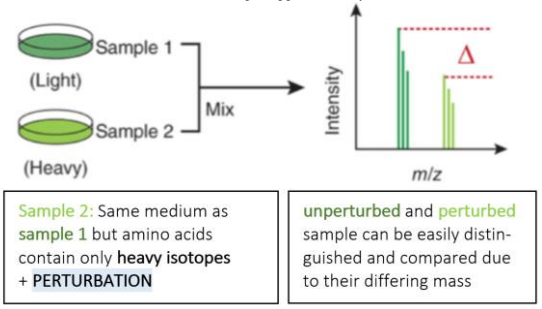
## SELECTIVE ISOTOPE LABELING

- PROBLEM: Complex sample preparation steps hinder reproducibility of results
- Selective isotope labeling: most reliable strategy for **comparing the abundance of proteins across samples**
- allows the simultaneous measurement of multiple samples and enables reliable comparison of protein abundance across samples

SILAC:

= stable isotope labeling by amino acids in cell culture

***GOAL**: measurement of how **perturbation** (mutation, change in growth conditions, the addition of a medicinal drug) changes the relative abundance of different proteins in cell*



# ERGÄNZUNGEN DER VORLESUNG

## MAJOR PRINCIPLES OF MOLECULAR BIOLOGY

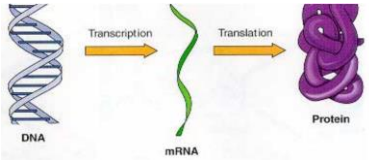
| | | |
|---|---|---|
| Mendel's Laws | 1865 | Mendelian inheritance |
| Avery | 1944 | DNA as transforming principle |
| Beadle & Tatum | 1941 | One gene – one protein – one function |
| Central Dogma (Crick) | 1958 | DNA -> RNA -> protein |
| Pauling | 1949 | Sickle cell anemia as a molecular disease: Mechanistically connecting genomic variability to phenotypic variability |

## DNA VS PROTEINS

- **Nucleic Acids**
- We sequence a synthetic complement to a template sequence
- The synthetic template can be extensively amplified
- We can sequence minute amounts of nucleic acids. E.g. single cells

- **Proteins**
- We directly sequence the proteins extracted from biological samples
- No amplification
- We **cannot** sequence minute amounts of protein e.g. from single cells
- It is a lot easier to sequence a gene than a protein

## CENTRAL DOGMA OF BIOLOGY




## PROTEINS

Crick (1958) „once (sequential) information has passed into protein it cannot get out again"

- Protein analysis is harder than DNA analysis
- Protein quantities cannot be predicted from DNA/RNA
- Proteins are frequently modified
- 20 different Amino acid

PROTEOME:

= the ensemble of all proteins of a cell or organism

- fission yeast; amazing complexity of the proteome
  - genome : 14 Mb, ~5000 genes
  - 42,000 mRNAs
  - Total number of proteins: 95 Mio per cell

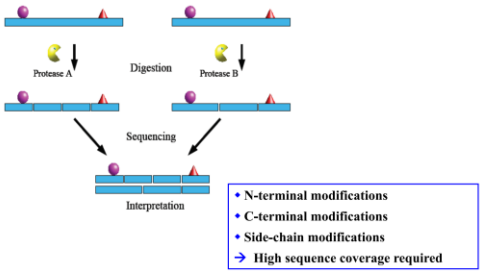
## POST-TRANSLATIONAL MODIFICATION

- Most eukaryotic proteins are modified
- There are hundreds of different protein modifications known, each with its specific functional and analytical characteristics
- Most modifications are relatively rare and inabundant in proteome digests and need to be enriched prior to MS
- Modified peptides frequently differ in the fragmentation patterns from non-modified peptides

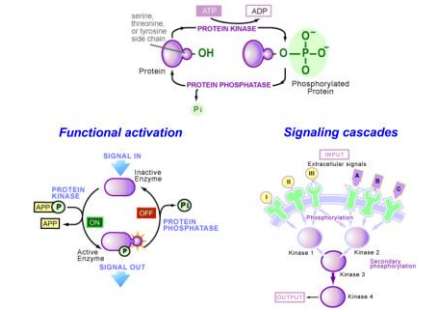
The determination of the precise localization of a modified residue in a peptide is frequently challenging

- Protein post-translational modifications
  - Modifications can be post- or co-translational
  - Many PTMs were discovered serendipitously during studies of individual proteins and biological processes
- Characterization of PTMs
  - Determine nature and site of modifications
  - Full PTM characterization is slow (sometimes not even possible), and requires isolation (enrichment) to obtain sufficient amounts
  - Conventional proteomics approaches are high throughput, but typically have little ability to characterize the full protein primary sequence

### Classical Analysis of PTMs



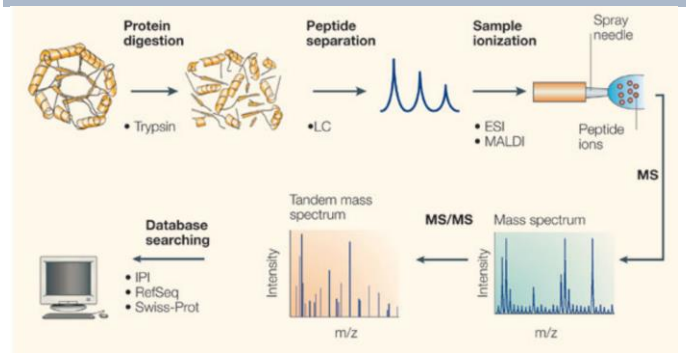

### Phosphorylation / Phosphoproteomics




### Modification Site Localization

- Typical proteomics search engines are not designed for localization of modification sites (does not influence scoring directly)
- Unambiguous assignment requires presence of **diagnostic fragment ions**
- High spectral quality required, but many peptides contain multiple possible modification sites
- Use of dedicated software tools or manual validation

## GENERIC PROTEOMICS METHOD VIA TANDEM MASS SPECTROMETRY



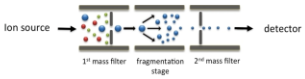
## Identifying Peptides from MS data

You have a sample that contains **one** of these two peptides: Peptide 1: **SGVNEGFFSR**

Peptide 2: **NEGFFSGVSR**

Could you use single-stage MS measurement of the peptide mass to determine, which of these two peptides is in the sample? **No, they have the same amino acid composition and hence the same mass. Monisotopic mass = 1099 Dalton (1080 (AA residue masses) + 18 ($H_2O$) + 1 (mass of proton) )**

What approach, that you have learned about in the hand-outs, will permit you to determine, which of the two peptides is present in the sample? **Fragmentation**

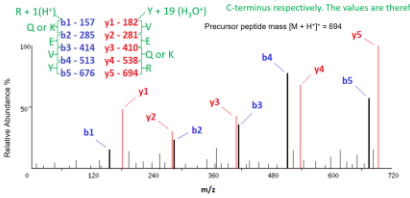



One of the fragments generated by fragmentation of the peptide along its backbone is detected in the second mass filter and has a mass of **487** Dalton. Can you use this information to determine, which of the peptides is in your sample? (use db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html to calculate the masses of all possible fragments from a peptide's sequence)

**Yes, fragmentation of Peptide 1 along its backbone creates a 487 Dalton fragment and Peptide 2 does not. The data supports that the peptide in the sample is Peptide 1.**

## Determining the complete sequence of a peptide from its b and y ion masses

The b1 and y1 ion masses are not those of the AA residue itself, but contain also a piece of the actual N- and C-terminus respectively. The values are therefore always off by 1 or 19 mass units respectively.




| 1-letter code | 3-letter code | Chemical formula | Monoisotopic Mass |
|---|---|---|---|
| A | Ala | $C_3H_5ON$ | 71 |
| R | Arg | $C_6H_{12}ON_4$ | 156 |
| N | Asn | $C_4H_6O_2N_2$ | 114 |
| D | Asp | $C_4H_5O_2N$ | 115 |
| C | Cys | $C_3H_5ONS$ | 103 |
| E | Glu | $C_5H_7O_2N$ | 129 |
| Q | Gln | $C_5H_8O_2N_2$ | 128 |
| G | Gly | $C_2H_3ON$ | 57 |
| H | His | $C_6H_7ON_3$ | 137 |
| I | Ile | $C_6H_{11}ON$ | 113 |
| L | Leu | $C_6H_{11}ON$ | 113 |
| K | Lys | $C_6H_{12}ON_2$ | 128 |
| M | Met | $C_5H_9ONS$ | 131 |
| F | Phe | $C_9H_9ON$ | 147 |
| P | Pro | $C_5H_7ON$ | 97 |
| S | Ser | $C_3H_5O_2N$ | 87 |
| T | Thr | $C_4H_7O_2N$ | 101 |
| W | Trp | $C_{11}H_{10}ON_2$ | 186 |
| Y | Tyr | $C_9H_9O_2N$ | 163 |
| V | Val | $C_5H_9ON$ | 99 |

The image above shows the fragmentation spectrum of a 5 AA-long peptide. Can you use the b- and y-ion masses together with the mono-isotopic masses of amino acids to determine the peptide's sequence?

Q
R K EVY

As you can see, if the complete set of b and y ions were obtained, the sequence of the peptide can be "read-out" both from the b- and the y-ion masses.

Why do the peaks have different height?

Not all of the peptide bonds are equally fragile and the efficiency with which they trigger a response of the detector varies also.

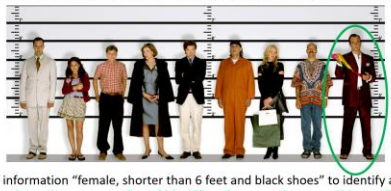
Why do the fragment ion peaks not have isotope subpeaks?

The first MS stage selected only one of the isotope peaks for fragmentation. The fragments therefor all come from an isotopically identical precursor pool.

## How much information do we need to uniquely identify an individual

If you had the following information about a person: taller than 6 feet, wears a suit, pants are not white. Could you uniquely identify that person? **No, there are almost 8 billion people on earth many of them are likely to fit this description.**

With the same information, could you identify the person from the "candidate pool" shown below?




Could you use the information "female, shorter than 6 feet and black shoes" to identify a unique individual?
**No, even though we have the same number of identifier, there are three possible candidates in the candidate pool.**

In general terms, what criteria determine the chance of uniquely identifying a candidate from a candidate pool with just partial information?
**The smaller the candidate pool relative to the theoretically possible pool, the less information is needed.**

## How good are the chances of uniquely identifying peptides from "incomplete" fragmentation spectra

The digestion of proteins with trypsin results in peptides that are ~10 AA long.

Considering only the 20 standard AA. How many different 10 AA-long peptides are theoretically possible based on pure combinatorics? **20 x 20 x 20 x 20 … = $20^{10}$ = 1.024 x $10^{13}$**

Would it be possible to uniquely identify each of these peptides, if all fragments are present in the fragmentation spectrum ? **If all AA had different masses yes. But some AA have identical masses (e.g. leu, ile or lys, gln)**

Considering that the human genome contains about 20'000 protein coding genes and that the average length of a protein is 300 AA. How many different peptides would be present in a trypsin digest of the complete human proteome.
**20'000 x 300 / 10 = 600'000 = 6 x $10^5$**

How does the number of peptides in the human proteome compare to the number of possible peptide sequences?
**The pool of AA contained in the human proteome is many orders of magnitude smaller than the pool of possible peptides.**

What does that mean for the completeness of data required to uniquely assign a measured fragmentation spectrum to a candidate peptide? **Just as in the example on the previous slide, it will often be possible to identify peptides uniquely, even, if just a few of their fragment ions are observed in the fragmentation spectrum.**

## Proteins are digested into Peptides prior to MS analysis

Imagine you have an isolated sample of human haemoglobin alpha-1 protein. The amino acid sequence of this protein is shown below. To prepare the samples for Mass-Spectrometry, you treat this sample with the enzyme trypsin to cleave it into peptides. Identify the peptides you will generate in this digestion reaction.

**SAQVKGHGKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAKFTPAVH ASLDKFLASVSTVLTSKYR**

Hints:

- When the sequence of a protein is written in single letter code, where is the N-terminus and the C-terminus of a protein?
- What is the sequence specificity of trypsin? (check hand out if necessary)

**SAQVK GHGK VADALTNAVAHVDDMPNALSALSDLHAHK LR LR VDPVNFK LLSHCLLVTLAAHLPAK FTPAVHASLDK FLASVSTVLTSK YR**

## Identifying proteins based on peptides

You work in a biotechnology company where you use an alcohol dehydrogenase from an external supplier. You are not sure, which form of the dehydrogenase they sell you, but you suspect that it is one of the four variants of this enzymes for which the peptide sequences are given below.

Using peptide fragmentation based MS, you successfully determined the sequence of four peptides found in the tryptic digest of your sample. a. **MSTAGK** b. **TVLTF** c. **AAVLWELK** d. **GAIFGGFK**

| Alcohol Dehydrogenase 1C (*Homo sapiens*) | Alcohol Dehydrogenase 1 (*Mus musculus*) | Alcohol Dehydrogenase Class-3 (*Mus musculus*) | Alcohol Dehydrogenase Class-3 (*Arabidopsis thaliana*) |
|---|---|---|---|
| MSTAGKVIKCKAAVLWELKKPFSIEEVEVAPPKAHEVRI<br>KMVAAGICRSDEHVVSGNLVTPLPVILGHEAAGIVESV<br>GEGVTTVKPGDKVIPLFTPQCGKCRICKNPESNYCLKN<br>DLGNPRGTLQDGTRRFTCSGKPIHHFVGVSTFSQYTVV<br>DENAVAKIDAASPLEKVCLIGCGFSTGYGSAVKVAKVTP<br>GSTCAVFGLGGVGLSVVMGCKAAGAARIIAVDINKDKF<br>AKAKELGATECINPQDYKKPIQEVLKEMTDGGVDFSFE<br>VIGRLDTMMASLLCCHEACGTSVIVGVPPDSQNLSINP<br>MLLLTGRTWKGAIFGGFKSKESVPKLVADFMAKKFSLD<br>ALITNILPFEKINEGFDLLRSGKSIRTVLTF | MSTAGKVIKCKAAVLWELHKPFTIEDIEVAPPKAHEVRI<br>KMVATGVCRSDDHVVSGTLVTPLPAVLGHEGAGIVES<br>VGEGVTCVKPGDKVIPLFSPQCGECRICKHPESNFCSRS<br>DLLMPRGTLREGTSRFSCKGKQIHNFISTSTFSQYTVVD<br>DIAVAKIDGASPLDKVCLIGCGFSTGYGSAVKVAKVTPG<br>STCAVFGLGGVGLSVIIGCKAAGAARIIAVDINKDKFAK<br>AKELGATECINPQDYSKPIQEVLQEMTDGGVDFSFEVI<br>GRLDTMTSALSCHACGVSVVVGVPPNAQNLSMNPM<br>MLLLGRTWVKGAIFGGFKSKDSVPKLVSDFMAKKFPL<br>DPLITHVLPFEKINEAFDLLRSGKSIRTVLTF | MANQVIRCKAAVAWEAGKPLSIEEIEVAPPKAHEVRIKI<br>LATAVCHTDAYTLSGADPEGCFPVILGHEGAGIVESVGE<br>GVTKLKAGDTVIPLYIPQCGECKFCLNPKTNLCQKIRVT<br>QGKGLMPDGTSRFTCKGKSVFHFMGTSTFSEYTVVAD<br>ISVAKIDPSAPLDKVCLLGCGISTGYGAAVNTAKVEPGS<br>TCAVFGLGGVGLAVIMGCKVAGASRIIGIDINKDKFAKA<br>KEFGASECISPQDFSKSIQEVLVEMTDGGVDYSFECIGN<br>VKVMRSALEAAHKGWGVSVVVGVAASGEEISTRPFQL<br>VTGRTWKGTAFGGWKSVESVPKLVSEYMSKKIKVDEF<br>VTGNLSFDGINQAFDLMHSGDSIRTVLKM | MATQSQVITCKAAVAYEPNKPLVIEDVQVAPPQAGEV<br>RIKILYTALCHTDAYTWSGKDPEGLFPCILGHEAAGIVES<br>VGEGVTEVQAGDHVIPCYQAECRECKFCKSGKTNLCG<br>KVRSATGVGIMMNDRKSRFSVNGKPIYHFMGTSTFSQ<br>YTVVHDVSVAKIDPTAPLDKVCLLGCGVPTGLGAVWN<br>TAKVEPGSNVAIFGLGTVGLAVAEGAKTAGASRIIGIDI<br>DSKKYETAKKFGVNEFVNPKDHDKPIQEVIVDLTDGGV<br>DYSFECIGNVSVMRAALECCHKGWGTSVIVGVAASGQ<br>EISTRPFQLVTGRVWKGTAFGGFKSRTQVPWLVEKYM<br>NKEIKVDEYITHNLTLGEINKAFDLLHEGTCLRCVLDTSK |

Based on the identified peptides' sequences, which of the enzymes do you think is in your sample? human alcohol dehydrogenase 1C.
Out of the four possible candidates it is the only protein that contains the peptide **AAVLWELK**
Bases on your data, can you be sure that neither of the other enzymes is present as well? **No, proving absence of a protein is not possible**

What properties must a peptide have so that it can be used to confirm the presence of aprotein in the sample?
**The peptide must be present in the protein that is to be identified but not in any of the other proteins that might also be in the sample, such a peptide is said to be proteotypic**

## ÜBUNGSTESTFRAGEN:

### PART 1)

*1. Discuss how (i) excitation wavelength, (ii) brightness of the fluorophore and (iii) pixel size on the CCD camera detector can influence the localization precision that you can obtain in PALM*

*2. A researcher performs a FRAP experiment on 'Stress Granules', cytoplasmic structures involved in RNA storage and degradation. He has labeled a 'Stress Granule component with GFP.*

*(i) Describe in the form of a bullet point list the principle steps of FRAP*

*(ii) The fluorescence intensity that the researcher measures in the Stress Granule at the end of the experiment is 50% from the initial value . What does this tell you about the GFP-labeled Stress Granule component? [5 points]*

### PART 2)

*Question 1*
*2nd and 3rd generation sequencing techniques have very different performance characteristics.*

a) *Please explain in a concise paragraph what the parameters: "read-length", "coverage", "throughput" and "error rate" describe.*
b) *Name one example each of a 2nd and a 3rd generation sequencing technique and give a rough value of their read-length and error rate.*
c) *Use your knowledge of the molecular process underlying these two sequencing techniques to explain the big differences in read-length, error rate and error type between these two techniques.*

*Question 2*
*Many of the biological DNA sequences investigated by sequencing (e.g. entire genes or even entire genomes) are much longer than the reads that can be obtained by any of the available sequencing technologies. Also, the error rates of the raw sequencing reads obtained by 2nd and 3rd generation sequencing technologies are very high relative to the frequency of real genetic variations that are investigated in most research projects. The questions below are about how short and error-prone primary sequencing data can be assembled into long and reliable sequences. In answering the questions please feel free to use diagrams or drawings to illustrate your point.*

**a)** *How does the shotgun approach use sequencing data of short, random DNA fragments to reconstruct the continuous sequence of the much longer unfragmented original DNA molecules. And, what determines the shortest read length that can, theoretically, be used to determine a long DNA sequence (e.g. a human genome) with the shotgun approach.*

**b)** *Explain how raw sequencing data with relatively high error rates can be assembled into final DNA sequences with very low error rates. In particular, explain how one distinguishes between a sequencing error (e.g. base call error) and an actual change in the DNA sequence (e.g. single nucleotide polymorphism).*

*Question 3*
*You are working as the sequencing specialist in a medical team that is treating an individual with a severe genetic disease. You are suspecting that the cause of the disease is a large scale chromosomal rearrangement.*

a) *Why may complete genome-resequencing with short-read, sequencing-by-synthesis technology not be the ideal technology to investigate a chromosomal rearrangement?*
b) *What other technology may provide a faster and more reliable way to detect and localize a chromosomal rearrangement?*

*Question 4*
*In the lab, you have just "cloned" a PCR product that is about 4000 bp long. That is to say you ligated this PCR product into a plasmid and have purified many copies of this plasmid. You now want to determine the sequence of the cloned PCR product.*

a) *What sequencing technique will you likely want to use and why?*
b) *Describe how you will design the sequencing primer?*
c) *Describe how you will continue with your work once you received your first sequencing results?*

*Question 5*
*The massspectrometers used for peptide-fragment-based protein massspectrometry use two mass-filter stages.*

a) *What type of ions are separated in these two mass-filter stages?*
b) *What takes place between these two mass-filter stages?*
c) *Why is it necessary to use two mass-filter stages in order to identify a peptide?*

*Question 6*
*Peptide-fragment-based protein massspectrometry can be used to analyze the presence of proteins in complex samples (e.g. extracts from cells or tissues).*

a) *Please describe what is the raw data that is recorded in a typical peptide-fragment-based protein massspectrometry experiment.*
b) *How will you use this this raw data to determine, which proteins were present in the sample? Explain the importance of proteotypic peptides in this context.*

*Question 7*
*A colleague of yours from a marine biology laboratory has discovered a new worm species that lives at the bottom of the Pacific Ocean. Remembering what you had told her about studying protein expression in human tissues via protein MS, the first thing she now wants to do is to analyze the proteome of this new organism with peptide-fragment-based protein massspectrometry.*

a) *Would the approach you are using to study the presence of proteins in human tissue samples work for the samples from the worm? Explain why.*
b) *What should your colleague analyze first before starting to analyze the proteome.*