

1. Plagiarism and Good Scientific Practice

1.1. Good Scientific Practice

Scientific misconduct is ...

- Plagiarism
- Fabrication, falsification, and removal of data/theft of data
- Missing the approval from other authors for submission of a publication manuscript
- Applying or using an authorship of honor
- Impeding/obstructing scientific work of colleagues

These rules are defined in the memorandum Safeguarding Good Scientific Practice, published by the Deutsche Forschungsgemeinschaft (DFG) in 1998. It was updated by the Guidelines for Safeguarding Good Scientific Practice published by the DGF in 2019.

Good scientific practice is warranted by ...

- Correct study planning and design
- Adequate methodology
- Correct performance
- Correct data documentation
- Correct evaluation of data
- Correct publication.

Correct scientific practice ...

- Requires correct dealing with your own data
- Requires correct dealing with data of others
- Requires correct data documentation in a specified protocol book
- Requires correct data publication and a correct publication process (including authorship regulations)
- Requires correct and fair interaction with colleagues (of the own group as well as of competing groups).

1.2. Plagiarism

Plagiarism is the act of presenting another's work, ideas or words as your own without proper citation. It is a serious offence under the University's regulations on cheating and will consequently be penalized. In extreme cases it can result in de-registration of the specific student.

Intentional plagiarism is ...

- Copying an essay or article from a journal, the internet, an online-source, or a database without quoting or giving credit to these originals.
- Cutting and pasting from more than one source to create a paper without quoting or giving credit.
- Borrowing words, sentences or ideas from other sources without giving credits.

Unintentional plagiarism is ...

- Paraphrasing poorly: changing a few words without changing the sentence structure of the original or changing the sentence structure but not the words.
- Quoting poorly: putting quotation marks around part of a quotation but not around all of it or putting quotation marks around a passage that is partly paraphrased and partly quoted.
- Citing poorly: omitting and occasional citation or citing inaccurately.

Common knowledge sentences – although they have once been described first – does not need to be cited.

Prevention of plagiarism

- Develop a topic based on previously written material but write something new and original in your own words,
- Rely on opinions of experts on a topic but improve upon those opinions.
- Follow a standard documentation method.

Plagiarism is a major problem in science and especially in the preparation of doctoral theses in the humanities and life sciences.

1.3. Research misconduct

Worldwide research misconduct is defined as fabrication, falsification or plagiarism in proposing, performing, or reviewing research results.

- Plagiarism: appropriation of another's ideas, processes, results, or words without giving proper credit
- Fabrication: making up results and recording or reporting them
- Falsification: manipulation of research materials, equipment, or processes, or changing or omitting results such that the research is not accurately represented in the record.

There are several historical cases of research misconduct, whereas the greatest German case was conducted by Hermann/Brach in cancer research in 1997.

1.4. Guidelines for keeping laboratory notebooks

How to keep a laboratory notebook

- Notebooks must be handbound and the pages must be numbered consecutively
- The laboratory notebook has to be kept chronologically
- Only permanent writing equipment may be used – pencils are not allowed.
- No pages shall be removed from the notebook
- To avoid documentation mistakes, documentation must be carried out promptly in the course of the experiment and immediately upon its conclusion.
- Each laboratory notebook is maintained by one and the same person in their own handwriting. This person is responsible for the proper keeping of notebook records.
- If more than one person is involved in the experiment, cross-references to the respective laboratory notebooks of all participants are necessary.

List of essential information that must be documented in the laboratory notebook

- Name, address, affiliation
- Table of contents
- Data, project title
- Designation to the experiment, expectations, references to repetitions
- Description of equipment used
- Materials used: nature and origin; if relevant batch number and storage
- If the experiments involve samples from patients this information must be included in pseudonymized form (patient number, sample number) without any direct reference to personal data
- The project coordinator/principal investigator is responsible for keeping an identification list to trace the data/samples back to the respective patient
- Detailed description of the experimental procedure, measurement protocols
- Original raw data of the samples and controls
- Calculated data
- Discussion, interpretation, and conclusion of results

→To ensure that the investigations are transparent and comprehensible for others, it is vitally important to provide detailed documentation that is both throughout and precise.

- If other data storage media is used, the laboratory notebook must contain precise details of the location of the original data as well as summaries and conclusions from the stored results.

- All documentations should regularly be dated and signed by the experimenter (and confirmed by a second person).

Where should the laboratory notebook be kept?

- They should not be publicly accessible.
- They belong to the institution in which the work was carried out. The institution is also responsible for archiving the protocol books and data for a period of at least 10 years.

2. Statistics

2.1. Terms in Statistics

Chemometrics = chemical discipline that uses mathematical, statistical, and other methods employing formal logic

- To design or select optimal measurement procedures and experiments
- To provide maximum relevant chemical information by analyzing chemical data

Linking elements between chemistry, mathematics, and hardware and software are used in sampling, quality control, experimental planning, to optimize and model processes and experiments, etc.

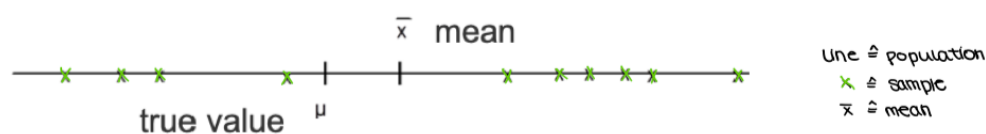
Sorting data:

- Advantages: appearance of accumulations
- Disadvantages: loss of trends

Sorted tables can be displayed as histograms (number plotted against measured values).

Definitions

- **Random variables** = analytically measured quantities
- **Population** = infinite amount of possible random variables e.g. all young adults
- **Sample** = certain set of random variables e.g. 50 men and 50 women
- **Distribution function** = probability of occurrence of a random variable around the **expected value** of a population
- **Expected value** = close to **true value μ**



Representation of a population always with an unbiased (representative) sample in order to reduce data collection time and costs. The probability of appearance of random variables is following a normal (gaussian) distribution.

Random variable

= numerical value out of measured value or result. The value can vary from trial to trial as the experiment is repeated. A random variable can have:

- Discrete values (**discrete random variable**)

e.g. Dice, display of digital watch

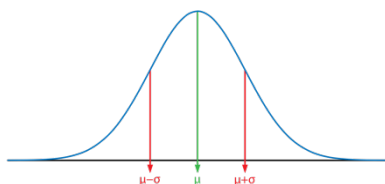
- Continuous value (**continuous random variable**) → continuous variation within one or more intervals
- **Continuous fluctuating values** around a certain value → probability of occurrence decreases with increasing distance, but value 0 is theoretically never reached
e.g. Analogue measuring instrument

A result is obtained after finishing an analysis. This analysis includes the completion of all measurements and all succeeding evaluation steps.

- **Mean value** ($\bar{x}$) is used for characterization of discrete distributions
- **Expected value** is used for characterization of discrete and continuous distributions
- **True value** μ does not need to be equal to measured value

Normal (Gaussian) Distribution

- Error between random value x_i and mean value $\bar{x}$: $e_i = x_i - \bar{x}$
- Error between random value x_i and true value μ : $\varepsilon_i = x_i - \mu = e_i + (\bar{x} - \mu)$
- Standard deviation $\text{sdv}(x) = s(x) = \sigma$:
 - One standard deviation away from mean counts 68.27 % of data
 - Two standard deviations count 95.45 %
 - Three standard deviations count 99.73 %
- Standard deviation of the mean $\text{sdv}(\bar{x}) = s(\bar{x})$



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Types of Errors

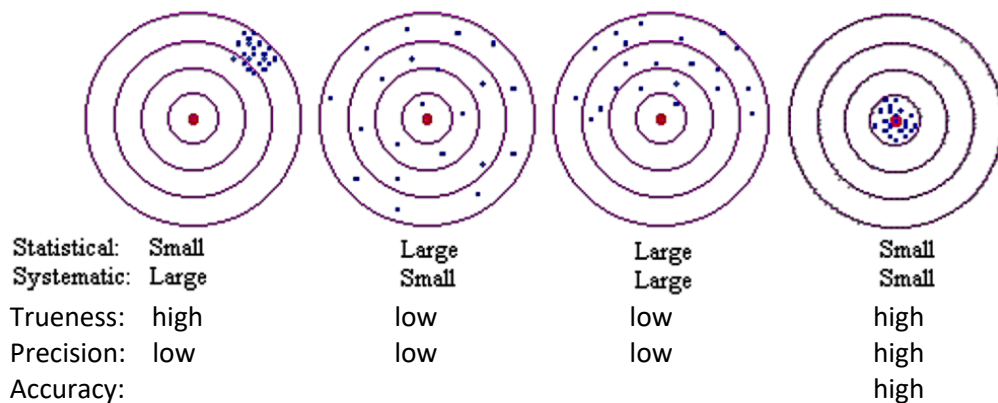
Results are distinguished from true value μ by an **error**. In contrast to the error, the **deviation** is the difference between a random variable and the mean within the same sample.

The **range** is defined as the difference between the largest and smallest value within a sample.

Two types of errors are important:

- **Statistical error = random** = precision (reproducibility) of an analytical process
 - Nonavoidable and cannot be characterized
- **Systematic errors = bias** = accuracy of an analytical process
 - Deviation of the results from the true value in the same sense

- Caused by e.g. defective instruments, wrong measurement methods and gross errors
- Can be observed by interlaboratory tests, but not by repeated measurements
- **Gross errors**
 - Should be avoided



Trueness, Precision, and Accuracy

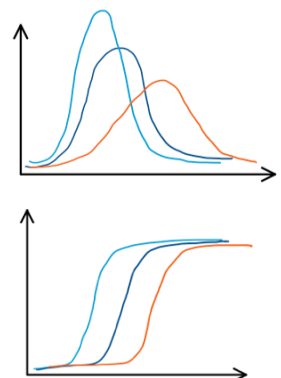
According to ISO 5725-1:

- **Trueness and precision** are used to describe the **accuracy** of a measurement
- **Trueness** refers to the closeness of the mean of the measurement result to the actual (true) value
- **Precision** refers to the closeness of agreement within individual results

Probability calculation

Obtained data are listed in a data matrix.

- **Frequency** (absolute number of a certain occurrence) and **relative frequency** (absolute number of a certain occurrence in relation to all occurrences) can be calculated out of the data matrix
- **Absolute frequency** = number of data
 - Requires **stochastically independent** measurement data
- **Equiprobability** = all occurrences have the same probability
- **Probability density function** = distribution curve = gives the frequency of equal occurrences
 - In case of discrete distribution also named probability function
- **Cumulative distribution function** = gives relative frequency
 - Symmetric (normal distribution) or skew (unimodal/bimodal/multimodal)



Convolution = mathematical operation on two functions producing a third function.

Binominal or Bernoulli distribution

= giving the probability that a certain event occurs x-times within n independently conducted experiments. Probability of a single event is p.

- Binomial distribution is most important discrete distribution
 - For $p = 0.5$ symmetrical, otherwise skew
- In every experiments only two types of results are possible, total probability of all results is equal to 1

$$f(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

n = independent experiments
 x = number of positive events
 p = probability (of a single event)

Poisson Distribution

- Can be applied to systems with a large number of possible occurrences ($k > 100$), each of which has a small probability ($p < 0.01$)
- Product of possible events and probability should be constant
- For large n and $p = 0.5$ normal distribution
- λ is equal to the expected value $E(x)$ of the discrete random variable X

$$p(k, \lambda) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

Example: Customers in a store

- store is entered every 10 seconds by one customer $\rightarrow \lambda = 6/\text{min}$
- $p_6(n)$ is probability for n customers to enter in the next minute

$$\rightarrow n = 1: P_6(1) = \frac{e^{-6} 6^1}{1!} = 1.49 \%$$

Derived statistics for a distribution

Estimator = any quantity calculated from the sample data which is used to give information about an unknown quantity in the population.

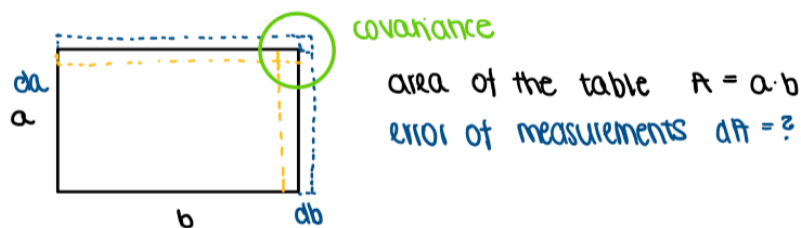
- E.g. sample mean is estimator for population mean
- Estimators can be distinguished from the true value by “hat” symbol $\hat{}$
 - σ = true population standard deviation
 - $\hat{\sigma}$ = estimated population standard deviation
- Point estimator: to calculate a single value (mean, variance)
 - Mean: $\hat{x} = \mu = \bar{x} = \frac{\sum x_i}{n}$

- Variance: $\sigma^2 = \frac{1}{n} \sum (x_i - \mu)^2$
- Variance of a single value: $var(x) = \frac{1}{n-1} \sum (x_i - \bar{x})^2$
- Variance of the mean: $var(\bar{x}) = \frac{var(x)}{\bar{x}}$
- Standard deviation: $sdv(x) = s(x) = \sqrt{var(x)}$
- Standard deviation of the mean: $sdv(\bar{x}) = s(\bar{x}) = \sqrt{var(\bar{x})}$
- Relative standard deviation: $resdv(x) = \frac{sdv(x)}{\bar{x}}$
- Interval estimator: to calculate several values (confidence interval)
 - Measurement uncertainty: $u_i = \text{statistical error} + \text{systematic error}$
 - Confidence interval of single value
 - Confidence interval of mean

Propagation of uncertainty

= effect of variables uncertainties (or errors) on the uncertainty of a function based on them

- E.g. uncertainties due to measurement limitations
- Uncertainty in one variable: can be estimated by Taylor propagation
 - In case of absolute errors approximation $y(x + \Delta x) - y(x) = \Delta y$
- Gaussian error propagation
 - Covariance $cov(x, y)$
 - Example: Covariance can be neglected, because errors of a and b are independent from each other



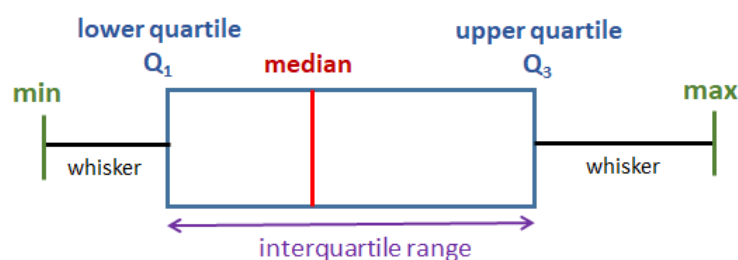
Robust statistics

= An estimation procedure or test procedure is called **robust** when it is not reacting sensitive to outliers (numerically distant from rest of the data).

- Mean value is influenced by fluctuations in the data set and by outliers

Example of robust statistics: Median statistics

- Median is used primarily for skewed distributions (summarized differently than arithmetic mean)
- Data presentation with Box plot possible
 - Shows median, quartiles and symmetry of distribution
 - $Q_1 = \frac{1}{4}$ of the data set, $Q_3 = \frac{3}{4}$ of the data set
 - Median = middle value of sorted data set
 - Whiskers = 1.5 times the IQR to identify outliers



2.2. Significance Tests

Analytical methods should be **free of systematic errors**. In order to decide whether the measured and standard values is significant or random, a significance test can be employed. Thereby, the truth of hypothesis (**Null hypothesis H_0**) is tested.

- One-sided hypothesis: $\alpha = 0.05$ use $q = 1 - \alpha$
- Two-sided hypothesis: $\alpha = 0.05$ use $q = 1 - \alpha/2$

Steps of testing

1. Write a null hypothesis H_0
2. Calculate the test by using the particular equation
3. Compare the calculated value with the tabulated value
4. Decision:
 - a. Calculated value < tabulated value $\rightarrow$ accept H_0
 - b. Calculated value > tabulated value $\rightarrow$ reject H_0

Students t-Test

- Corrects the finite number of data values to a big number of the normal distribution

- Dependent on the degree of freedom
- Tabulated: degree of freedom against P
- Comparison of an experimental mean $\bar{x}$ with a known value μ

$$t = \frac{\text{error of the mean}}{\text{standard error of the mean}} = \frac{|\bar{x} - \mu|}{\frac{sdv(x)}{\sqrt{n}}}$$

F-Test

- For the comparison of standard deviations
- Number of degree of freedom is n-1
- Requirement: populations from which the samples are taken are normal distributed

$$F = \frac{s_1^2}{s_2^2} \quad (s_1^2 > s_2^2)$$

- Test forms:
 - Test whether method 1 is more precise than method 2 (one-sided test)
 - Test whether method 1 and 2 differ in their precision (two-sided test)
- F-values are tabulated: Table for $\alpha = \dots$, whereas degrees of freedom of numerator and denominator are plotted against each other
 - $F_{\text{calc}} > F_{\text{tab}} \rightarrow H_0$ rejected
 - Variances inhomogeneous
 - No decision possible $\rightarrow$ Systematic error
 - Probability of error different
 - Population cannot be pooled
 - $F_{\text{calc}} < F_{\text{tab}} \rightarrow H_0$ accepted
 - Variances only randomly differ from each other
 - Population can be pooled

Paired $\hat{t}$ -Test

- It is tested whether the difference in the means is caused by random error
 $\rightarrow$ Requirement: F-Test in advance
- Standard deviation of the difference of the means is calculated by propagation of uncertainty
- If the differences $|\bar{x}_a - \bar{x}_b|$ are random variables follow t-distribution
- To test whether n paired results are drawn from the same population, calculate the statistical value $\hat{t}$

$$\hat{t} = \frac{|\bar{x}_a - \bar{x}_b|}{\frac{sdv(x)}{\sqrt{n_a + n_b}}} \sqrt{\frac{n_a n_b}{n_a + n_b}}$$

- $\hat{t}_{calc} > t_{tab}$ → significant difference between the two means → systematic error

χ^2 -Test

- Often used for the validation of crossbreed experiments
→ Verification with a certain probability whether a result (which is deflected from the ideal values) is corresponding with the theoretically expected experiments

Example:

- Mendel's experiments on plant hybridization

Kolmogorov-Smirnov Test

= Test whether the data are following a normal distribution

Example:

- Concentration of a component A is determined
 - Sorting values in ascending order
 - Z-transformation to find the area under the standardized normal distribution
 - Calculation of area distances → should be equal for normal distribution
 - Difference between calculated area under the standardized normal distribution and the ideal distances is calculated
 - Comparison with the tabulated value:
calculated value > tabulated value → significant difference from normal distribution

2.3. Analysis of Variance (ANOVA)

= describes the possibility to compare different random samples

- Random samples: from different methods or measured at different times
- Basis is a set of data consisting of columns and rows
 - Columns: results of different methods or times
 - Rows: different measurements of the random samples
- Analysis of variance until now:
 - t-Test: means are compared by variance analysis
 - F-Test: variances are compared

Heartley's test

- Ratio of largest and smallest variance

$$\hat{F}_{max} = \frac{\text{largest variance}}{\text{smallest variance}}$$

- Parameters:
 - Number of k groups
 - Number of degrees of freedom (n-1) for different groups
- Calculated value > tabulated value → Samples are not homogeneous
- Table shows degrees of freedom against number of groups

Cochran's test

- Equality of several variances: Variances of groups are significantly larger than the remaining

$$\hat{G}_{max} = \frac{s_{max}^2}{s_1^2 + s_2^2 + s_3^2 + \dots}$$

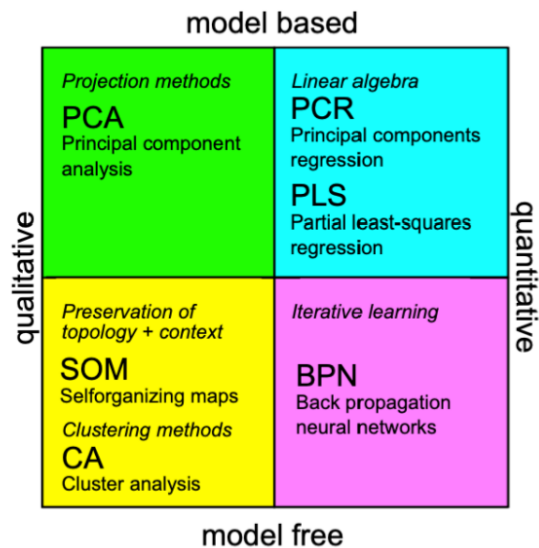
One-Way ANOVA

- Several determining factors included, observations are investigated more systematically
 - Comparison of more than two means, while considering that events are not individual and isolated → Total probabilities have to be considered
 - Requirement: Normal distributed data
 - Observed total variance is fragmented in parts, which are related to the same roots/attributes:
 - Random influence
 - Influence which has to be investigated
 - Variance analysis allows to identify:
 - The influence of the single variation for each measure or method respectively
 - The influence of the measurement series on the total variance
- Good to consider which experimental conditions lead to optimal results.
- Instead of many t-tests for comparison of the means of several experimental series, methods (M) or random samples (groups), just one F-Test between two variances is needed
 - Variances are
 - Summation of the differences between the means of the several groups (columns) and the total mean
 - Sum of squares of the single value from the mean of the group

- Squared sum of the deviation of the column mean (average over the row) from the total mean divided by the sum over all deviations from the column mean
- If groups originate from the same population, ratio will have nearly the same value
- Calculated value $\hat{F} >$ Tabulated value of F-distribution $F \rightarrow$ Rejection of H_0
- $n = RM$ if random sample size is equal for all methods (n is total number of all values of the overall M columns)
 - Standard deviation of single values around means of groups
 - Standard deviation of means of the groups around total mean
- Random sample error = mean squares over all groups
- Experimental error = Mean squares within a group

3. Multivariate Data Analysis

3.1. Introduction and Basic Aspects



Modelling

Hard modelling = number of variables or parameters is fixed

Soft modelling = number of variables or parameters not fixed, determined after running the regression analysis

- Data evaluation is either univariate or multivariate, depending on the number of variables
- “Multiple Regression” in context of classical algorithms
- Relationships can be solved either
 - Classical with a direct solution $y = f(x)$
 - With an inverse approach $x = g(y)$

Term: Variate

- **Univariate** = 1 independent, 1 dependent variable
 → e.g. measurement of one wavelength at different points in time

$$\mathbf{x}, \mathbf{y} \quad o = \text{Number of objects}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_o \end{pmatrix} \quad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_o \end{pmatrix}$$

- **Multivariate** = 1 independent variable (component), multiple dependent variables
→ e.g. spectral measurement
- **Multiple multivariate** = multiple independent variables (components), multiple dependent variables
→ e.g. Spectral measurements

$$\begin{array}{l} \mathbf{x}, \mathbf{Y} \quad v = \text{Number of variables} \\ \quad \quad \quad \text{e. g. number of wavelengths} \\ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_o \end{pmatrix} \quad \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1v} \\ y_{21} & y_{22} & \cdots & y_{2v} \\ \vdots & \vdots & \ddots & \vdots \\ y_{o1} & y_{o2} & \cdots & y_{ov} \end{pmatrix} \\ \\ \mathbf{X}, \mathbf{Y} \quad k = \text{Number of components} \\ \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{o1} & x_{o2} & \cdots & x_{ok} \end{pmatrix} \quad \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1v} \\ y_{21} & y_{22} & \cdots & y_{2v} \\ \vdots & \vdots & \ddots & \vdots \\ y_{o1} & y_{o2} & \cdots & y_{ov} \end{pmatrix} \end{array}$$

Data Analysis

- Univariate Regression: All measured values are depending on **one variable** only
- Multivariate Regression: more complex relationships can be described
 - Correlation of different kind of variables can be investigated simultaneously
 - Data set can be evaluated at the same time
 - Intention is to classify data set and to identify unknown properties

Requirement: Methods used are operating on the principle “soft modelling” (**parameterized**) or even model free (**non-parameterized**).

Aim of multivariate data analysis is to reduce the data set → data reduction

- Objects = samples, chromatograms, operating conditions of a plant, ...
- Pattern = contains the whole range of measurement results for one object
→ Pattern is a set of features selected for the description of the objects
- Features = variables

Linear Methods

- Interpretation and inspection of analytical data
 - Graphical representation
 - Correlation and identification of significant variables
- Evaluation of analytical data
 - Classification: Cluster analysis, PCA
 - Regression: univariate, MLR, PCR
- A data matrix $X_{(n, p)}$ contains objects in the rows and features in the columns

Polar plots can give insight in complex data.

3.2. Cluster analysis

= Method which groups objects of a data set into homogeneous classes (**clusters**)

- Objects with similar variable patterns are grouped into the same classes

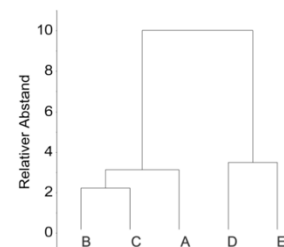
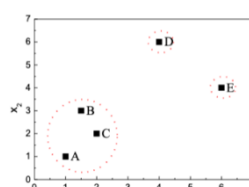
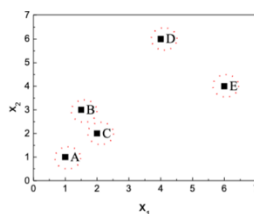
- Groups are unknown beforehand → automatic/unsupervised classification
 - Look for objects which are close to each other in the variable space
 - Calculate distance dimensions in order to group the objects
 - Develop a **dendrogram**

Hierarchical, agglomerative cluster analysis:

- Each element (measured value) is considered as a separate cluster
- Separate clusters are successively merged together to larger clusters beginning with the closest clusters (non-reversible)

Example:

1. Plot objects in the variable space
2. Look for groups/clusters between the objects
 - a. Every object is a cluster with size 1
 - b. Compare distances of the individual clusters with each other and look for smallest distance (Euclidean distance)
 - c. Group two closest objects to a new cluster
 - d. Repeat step b. and c. until all objects are grouped in a cluster
3. Plot of successive objects in a dendrogram
 - a. Vertical axis = relative distance between two objects (similarity)

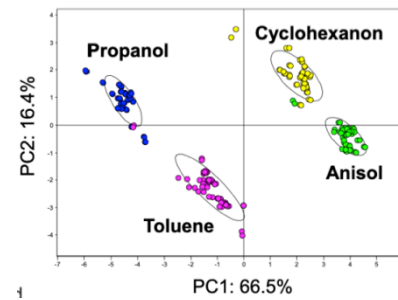


Other distance dimensions, such as Minkowski distance, City Block/Manhattan distance, Mahalanobis distance can also be used. → Disadvantage: All distance dimensions are depending on the used dimensional/measuring units.

3.3. Principal Component Analysis (PCA)

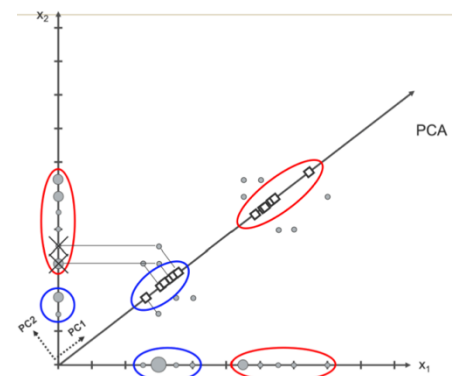
= Represents a 2D view on data, in which

- The maximum of the variance is contained
- While preserving the distances between the objects
- Often amount of one principal component is printed on axis label
- Principal components are sorted by their information content



Procedure:

- One dimensional mapping of values in two independent graphs (x_1 and x_2)
- Two dimensional mapping of values against each other (x-y graph)
 - New axis in direction of biggest variance (PC1)
 - Perpendicular to PC1 is PC2 axis



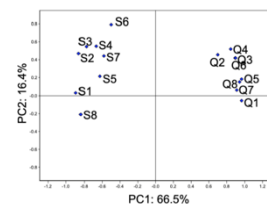
→ Method to separate relevant information from the noise.

In PCA, a data matrix $\mathbf{X}$ is decomposed to:

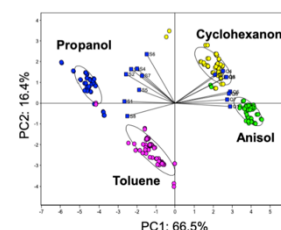
$$\mathbf{X} = \mathbf{USV}^T \text{ with } \mathbf{S} = \text{diag}(\sqrt{\lambda_i})$$

with $\mathbf{U}$ containing Eigenvectors of $\mathbf{XX}^T$, $\mathbf{V}$ containing Eigenvectors of $\mathbf{X}^T\mathbf{X}$, and $\mathbf{S}$ as a diagonal matrix containing the square roots of Eigenvalues λ .

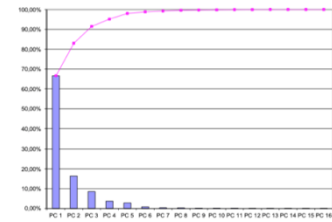
A **Loading Plot** shows the influence of each feature onto the principal components. Features not contributing will be near the center. Features with similar information content will be represented next to each other.



In a **Biplot**, scores and loadings are shown. By comparing the directions of the loadings with the position of the objects it is possible to draw conclusions on the suitability of individual features.



The **Scree plot** shows the amount of information (variance) represented by each principal components (bar). It is also possible to show information content covered by the first x components (line). Information content of a principal component is the size of its Eigenvalue, scaled in a way that the sum of all Eigenvalues is 1. Scree plot helps to identify number of components, necessary to explain data.



PCA can be used for:

- Fast inspections of the data (outliers, drift, main grouping)
- Unbiased representation of data
- Identification of feature structure (similarities, influence on objects)
- Reduction of dimensionality of the data
- Removing of noise
- Working with small datasets

Limitations:

- Unsupervised method → Does not take any knowledge into consideration
- PCA will not unveil desired information if high amount of noise
- PCA does not identify optimum feature composition for classification or regression tasks
- PCA requires careful weighting of input variables
- PCA plots may be misleading when the dimensionality of problem is high

4. Principle Component Analysis (PCR)

Regression: Basic Considerations

- Besides X information (data matrix) now also Y information (independent variable)
- Concept of independent test data: Data matrix to be split for model building (calibration) and independent test of the model (validation)

Multivariate Linear Regression (MLR)

Basic concept:

- Prediction of target values y from x , using a multi-dimensional Least Squares Fit
- Quantitative relationship between X and Y
- Sum of squared prediction errors of the training set is minimized

Mathematics:

- Relation: $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$
With X: matrix with independent variables, Y: matrix with dependent variables for objects, B: matrix with regression parameters, E: matrix with residuals (differences between predicted and true data)
- To reduce prediction error with MLR, either increase of number of objects or decrease of number of variables/features

Advantages and Limitations:

- Advantage: MLR is good for regression of data with no correlation within the X-variables
- Limitations: Correlation is practically always present in analytical data and especially chemical sensor data

Principle Component Regression (PCR)

- Extends MLR by introducing a PCA to calculate a new set of variables (PC), which are then used as input for a MLR
- Number of PCs is an important parameter
- If all PCs are used, the result is identical to MLR

→Weighting of input variables is an issue since it affects the principal components

- With PCR choice to disregard PCs that provide only little or noisily information.

Advantages and Limitations:

- Use of PCR for regression on data, where

- A limiting set of calibration data is available
 - The desired information can be found (visually) in the scores plot
- Limitations:
 - Model building does not optimize for the desired information
 - Nonlinear data may lead to problems

5. Bioanalysis: Separation and Diagnosis

Bioanalysis comprises DNA/RNA, proteins and their modifications (phosphorylation, glycosylation, ...), metabolites, lipids and interactions between these classes. Bioanalysis is used for systems biology (cell culture, mouse models), genomics, proteomics, metabolomics, quality control in biopharma, biomarker research and validation, clinical chemistry, biotechnology, forensics, ... It includes preanalytical methods (study design, sample preparation, extraction, centrifugation, purification), analytical techniques (chromatography, electrophoresis, ...) and data evaluation (preprocessing, ...)

Separation Techniques and Principles

PHASES	SEPARATION PRINCIPLE/EQUILIBRIUM	METHOD
GAS-LIQUID	Solubility differences for gases in liquids	Gas chromatography
GAS-SOLID	Differences in absorption of gases on solids	Gas chromatography
LIQUID-LIQUID	Differences in distribution	Liquid chromatography
LIQUID-SOLID	Differences in absorption of dissolved solutes onto solids	Liquid chromatography
	Differences in solubility in liquids	
EXTERNAL FIELD	SEPARATION PRINCIPLE	METHOD
ELECTRIC	Different mobility in electric field	Electrophoresis

5.1. Chromatography

Idea of Chromatography

- Repetitive equilibration at front and end of the analyte zone ("band")
- Movement through column by fresh mobile phase
- Retention of analyte in/on stationary phase depending on interaction
- Detection at column end to obtain a chromatogram
- Elution time: identity
- Band height or area: quantity
- Two phases: mobile and stationary phase, which do not mix
 - Sample is dissolved in mobile phase
 - Analytes are transported over the stationary phase

- Equilibria between mobile and stationary phase:
 - Distribution/partitioning
 - Or adsorption onto solid stationary phase
- Repetitive equilibration along the chromatographic column

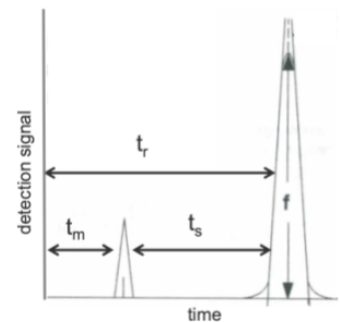
Types of equilibria:

- Adsorption (solid stationary phase) → adsorption chromatography
- Distribution (concentration) → normal/reverse phase liquid chromatography
- Complexation by e.g. cyclodextrins → chiral chromatography
- Ligand binding (antibodies) → ligand exchange chromatography
- Ionic interaction → anion/cation exchange

Strong interactions – strong retention – late elution

Separation Process

- Not-retarded substance needs time t_m to be transported through column (“void volume”)
- t_m is a measure of the mean velocity of the mobile phase
- all substances stay in mobile phase for time t_m
- Retention time t_r for analyte: time from injection to detection
- Analyte is retained and thus stays on the stationary phase for the time t_s



Phase equilibria – Retention factor

Distribution constant K_c cannot be measured directly → retention time and factor are used

1) Phase equilibrium with retention factor k (dimensionless)

$$k = \frac{\text{time spent in stationary phase}}{\text{time spent in mobile phase}} = \frac{\text{amount of substance in volume of stationary phase}}{\text{amount of substance in volume of mobile phase}}$$

$$k = \frac{t_r - t_m}{t_m} = \frac{n_s}{n_m} = \frac{c_s V_s}{c_m V_m}$$

2) Separation selectivity α

2 Substances with k_1 and k_2 are distributed between the two phases m and s

→ Selectivity:
$$\alpha = \frac{k_2}{k_1}$$

→ The highest α the easier is the separation of the two compounds.

Separation in the column

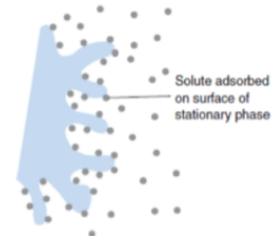
Concentration distribution (Gaussian peak shape) flattens with increasing number of equilibration steps.

Classification of different chromatography modes *via* phases.

Chromatographic separation mechanisms

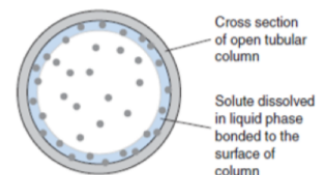
Adsorption:

- Describable via adsorption isotherms
- Mostly used for gas analysis in gas chromatography



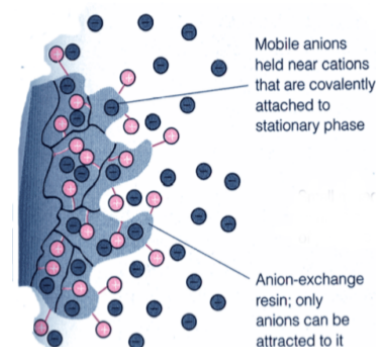
Distribution/Partitioning:

- Liquid film or stationary phase bound onto solid support (mostly silica gel)
- Distribution/partitioning between two phases
- Normal phase NPLC: polar stationary phase, nonpolar mobile phase (pentane, hexane)
- Reverse phase RPLC: nonpolar stationary phase (silanol-groups of silica gel derivatized with dichlorosilane and nonpolar alkyl chains) → mostly C₄, C₁₈, polar mobile phase (water, acetonitrile, methanol)
 - Retention based on polarity: high retention for low polarity
 - RPLC with C₁₈ stationary phase for most applications today, e.g. proteomics, metabolomics, environmental analysis



Ion exchange/ionic interaction:

- High retention for ions of higher charge, decreased hydrated radius and increased polarizability



5.2. Mass Spectrometry (MS)

MS is used to analyze atoms, molecules or their fragments.

Working principle:

- Gaseous compounds from liquid phase are ionized

- Ions are accelerated in electric field
- Separation according to mass to charge ratio m/z

→ Mass spectrum shows number of detected ions for each m/z

Different ionization procedures can be applied:

- Electron spray ionization (ESI): transfer of ionic species into gas phase
 - Dispersal of fine spray of charge droplets
 - Solvent evaporation
 - Ion ejection from highly charged droplets
- Quadrupole time-of-flight

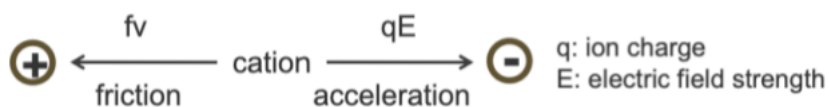
Applications of LC-MS:

- Metabolite profiling of distinct drugs, e.g. buspirone
- Proteomics

5.3. Electrophoresis

= Movement of charged particles in an electric field

An ion of charge q experiences acceleration force qE in electric field E and friction force fv with friction coefficient f :



→ Hückel equation:
$$v = \frac{qE}{6\pi\eta r} = \mu E$$

With μ = electrophoretic mobility, r = radius, η = viscosity

Basic model:

- Small sample plug in homogeneous electrolyte solution and application of voltage
- After some time, full separation of cationic analytes
- Separation dependent on electrophoretic mobility μ

Factors influencing the separation:

- Analyte properties
 - Analyte charge
 - Analyte size
 - Shape
 - Diffusion coefficient

- External parameters
 - pH of electrolyte
 - viscosity
 - applied voltage
 - capillary length

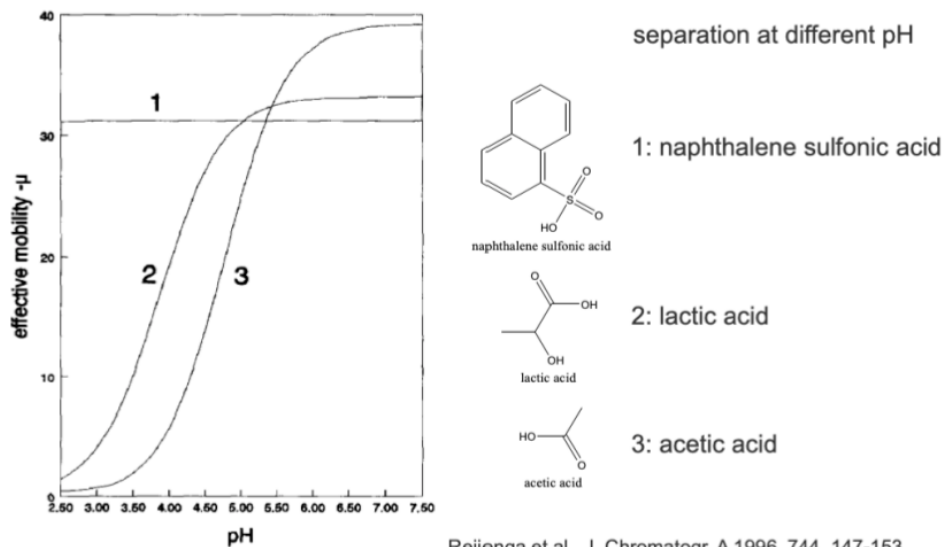
→ Fast migration of highly charged, small ions.

Applications:

- Separation of aromatic amino acids → separation according to charge, therefore dependent on number of amine functional groups
- Separation of homologues of aliphatic fatty acids
 - Fatty acids only differ in length of alkyl chain and thus in size, but not in charge

pH-dependence of mobility of molecules:

- Sulfonic acids have very low pK_a and are always fully charged in aqueous solution, therefore: electrophoretic mobility independent on pH-value
- Lactic and acetic acid: plot shows their titration curve
 - At low pH protonated and not charged
 - At $pH = pK_s$ charged by 50 %
 - Fully deprotonated with charge number = 1 at high pH



Modes of electro-migrative separation techniques: Capillary electrophoresis, CE-SDS, DNA sequencing, 2DE, electro-migrative separations, capillary isoelectric focusing, natice and SDS-PAGE.

Example: Ethyl sulfate as alcohol marker

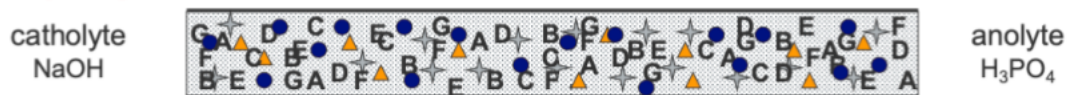
- Ethyl sulfate is strong acid ($pK_a < -3$), fully charged in aqueous buffers
- Separation at very low pH = 2.2
 - Only ethyl sulfate and inorganic/small organic ions are still charged
- High selectivity as most matrix is positively charged or neutral
- Polarity: inlet negative, outlet positive, -10 kV applied voltage
- Ethyl sulfate without chromophore → detection via indirect detection (replacement of absorbing buffer component)

Isoelectric focusing

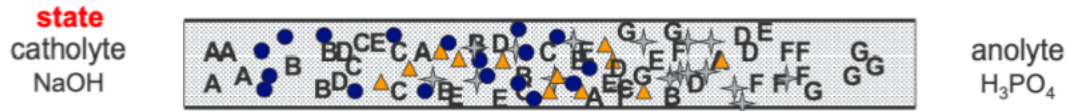
= Technique for separating different molecules by differences in their isoelectric point pI (negative and positive charged balanced at isoelectric point)

General procedure:

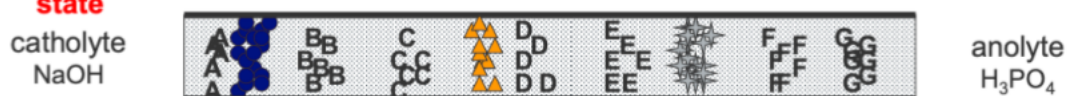
sample injection



transient state



stationary state



preconcentration and focusing

A,B,C,D,E,F,G: ampholytes

•, ▲, ☆: analytes, markers

Sample and Buffers:

- Sample in ampholyte mixture filling the whole capillary
- Inlet: catholyte – strong base (mostly NaOH) – high pH
- Outlet: anolyte – strong acid (mostly H_3PO_4) – low pH
- Or: immobilized pI gradients

Transient state:

- Migration of OH^- and H^+ into/through capillary
- Acid-base reaction with ampholytes/analytes

- Migration of all components until their pI is reached → focusing

Stationary state:

- After ca. 20 min
- No further migration
- Self-sharpening of bands

Mobilization:

- Pressure of pH (exchange of catholyte NaOH by acetic acid)

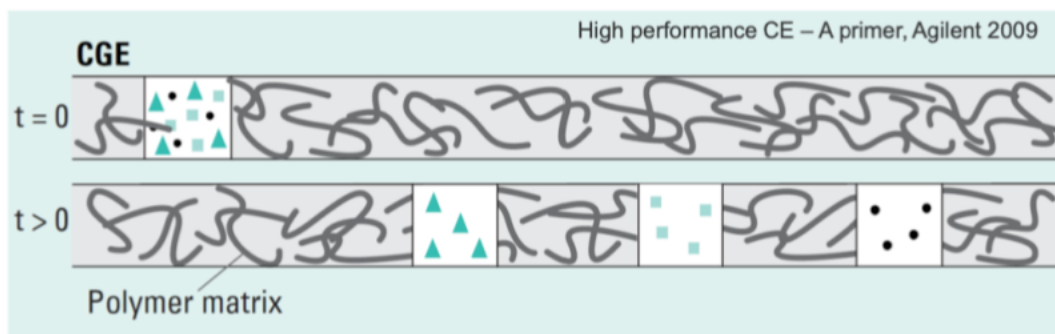
OR whole column detection (with laser-induced fluorescence) or second dimension PAGE in 2DE.

Size sieving electrophoresis

- Can be compared to normal gel electrophoresis → addition of a sieving matrix for size-based separation

Capillary size sieving electrophoresis (CSE):

- Applicable for separation of analytes with identical mass to charge ratio
e.g. DNA with every base being single phosphorylated or proteins being saturated with charged surfactant SDS
- Used for
 - Size-based separation of proteins (SDS-PAGE, CE-SDS)
 - Size-based separation of DNA
 - DNA sequencing
- Fields of application: forensics, clinical diagnostics, protein characterization, DNA sequencing, identification of food

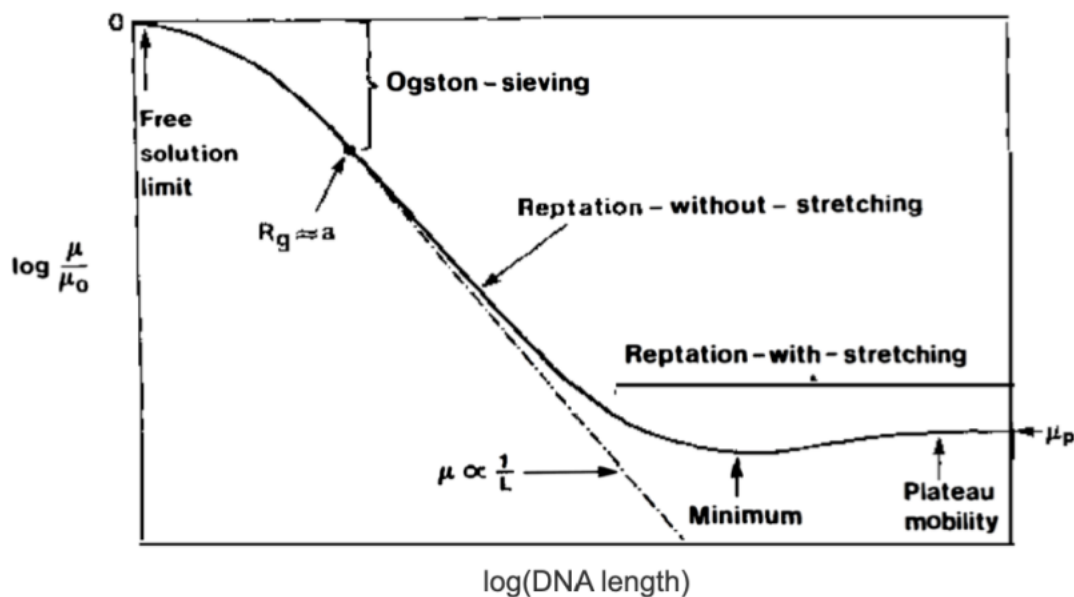


Sieving matrices:

- Gels:
 - For slab gel format

- Mostly cross-linked polyacrylamides
- Pore size via relative concentrations, e.g. 7 % gel with large pores, 12 % small pores
- Physical “gels”: Polymer network, entangled polymers (“intertwined chaines”)
 - Only for capillary format
 - Made of hydrophilic polymers in buffer
 - Forming a network from polymer molecules
 - Concentration gives mesh size of the sieve

Model of size sieving for DNA:



- Free migration: mobility is not size dependent but equal to free migration in solution
- Ogston model: average pore size is about the hydrodynamic radius
 - Classical sieving effect: DNA migrates similar to spherical particles through the pores
 - Exponential dependence on molecular weight
- Reptation regime: analytes are larger than pores and cannot simply penetrate
 - DNA has to stretch out and move snake-like through pores
 - Mobility negatively correlated with molecular weight
- Biased reptation regime
- Plateau mobility: Molecules orient in field direction, no more size-based separation

DNA color sequencing:

- Procedure:
 - DNA splicing into two strands

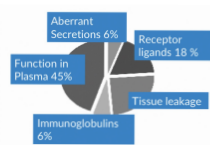
- Attachment of small DNA primers
- Enzyme DNA polymerase synthesizes complementary strand using artificial nucleotides
- Addition of fluorescence-labeled blocker nucleotides (dideoxy sugars)
- Analysis of new DNA strands
- Example:
 - Rhodamine colorants for blocker nucleotides with low spectral overlap → color code for sequencing
 - Genomic sequencing analysis on chips allows analysis of large sample cohorts
 - 2DE: IEF plus SDS-P

Definitions

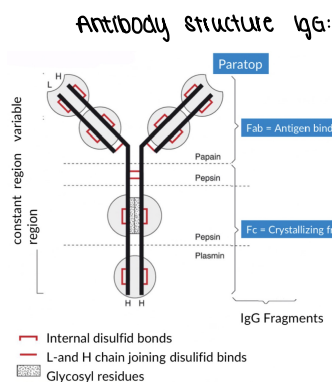
- Affinity = describes strength of antibody-antigen binding
- Antigen = molecule which is capable of triggering antibody production, antigen $\hat{=}$ analyte
- Antibodies (Immunoglobulins): Plasma proteins which are generated by vertebrates to neutralize xenobiotics, bacteria & viruses
Antibodies bind specifically towards their antigen. Several classes (IgG, IgM) & subclasses exist
↳ IgG antibodies are the major antibody class which is employed in immunoassays
- Crossreactivity = Binding of an antibody to molecules which comprise similar or identical structures (epitopes)
- Epitope = Part of antigen which interacts with the antibody
- Immunoassay = Every bioanalytical test in which an antibody-antigen reaction is employed for molecule detection
- Paratope = Part of antibody which interacts with the antigen
- Specificity = capability of an antibody to discriminate between its antigen & other molecules

Introduction

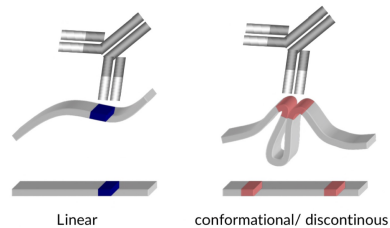
- Plasma Biomarker → Challenge: blood serum protein concentrations are in a very wide range
 - ↳ classical plasma proteins mg/mL - µg/mL
 - ↳ tissue leakage ng/mL
 - ↳ interleukins, etc. ng/mL - pg/mL
- Protein assays in diagnostics:
 - total number of FDA approved non-redundant protein assays: 109
 - novel approvals per year 1.5
 - additional 98 protein assays on the menu of large clinical labs
 - about 1% of the human proteome



Antibodies and other capture molecules

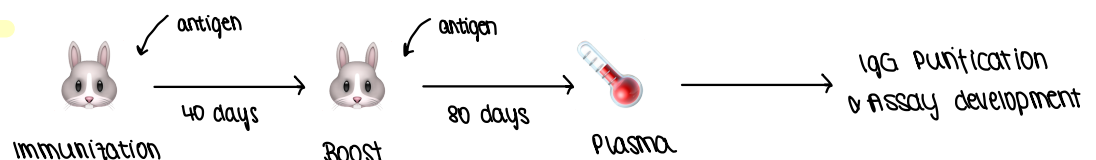


Linear vs. conformational epitope



- Antibody - Antigen Interactions: different binding modes:
 - I. Hapten Binders e.g. small molecules → deep pockets
 - II. Oligomer Binders e.g. peptides & DNA → cleft
 - III. Protein Binders e.g. proteins → large contact area
- Affinity: Formation of antibody-antigen complex $\rightarrow A + B \xrightleftharpoons[k_d]{k_a} AB$ with $K_D = \frac{[A] \cdot [B]}{[AB]} = \frac{k_d}{k_a} = \frac{1}{K_A}$
 K_D : Dissociation constant
 K_A : Association constant

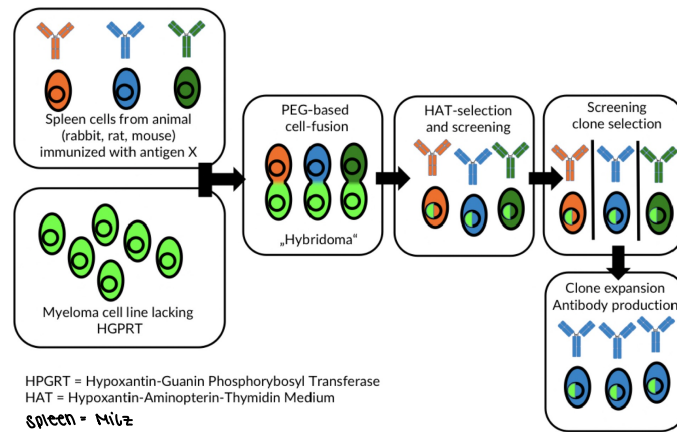
Classical Immunization



↳ animals mono- & polyclonal rabbit, mouse, rat

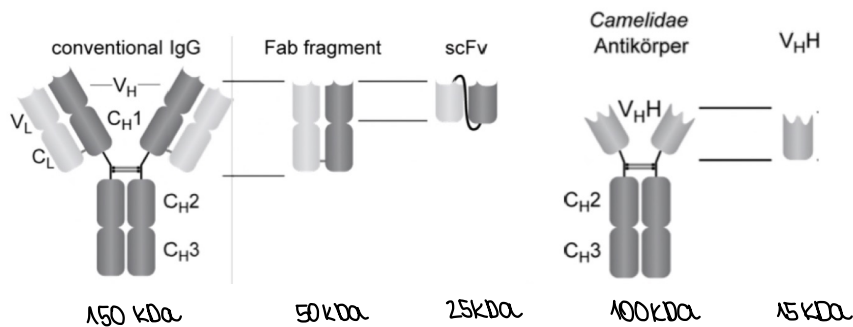
↳ animals polyclonal: donkey, guinea pig, chicken, goat

Antibody generation: monoclonal antibodies



Recombinant antibodies

- conventional IgG antibody
 - ↳ Fab fragments (antigen-binding region)
 - ↳ Single-chain variable fragments (fusion protein of the variable regions of heavy & light chains)
- Camelidae antibodies without light chains
 - ↳ V_HH antibody (nanobody (antigen binding fragment of heavy chain only))



Recombinant antibody generation:

1. Isolate DNA encoding V_L & V_H domains
2. Construct fusion of V regions & phage coat protein
3. Enrich antigen-specific phages (several rounds) & maturation process

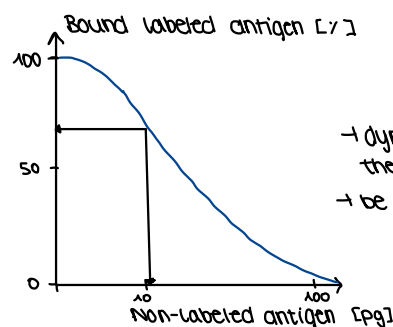
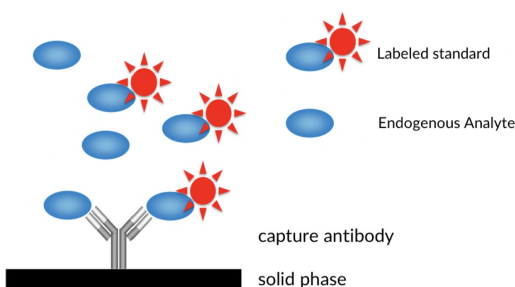
Other capture molecules: Ankyrin (DARPin), Protein A (Affibody), Lipocalin (Anticalin), Aptamer (Somamer)

- ↳ they differ in development time, antigens, yield & work load

Competitive immunoassay

Examples: Radioimmunoassay (RIA), Competitive ELISA, Lateral flow assay

Principle Radioimmunoassay



↳ dynamic range < 100%, therefore repetitions are necessary
↳ be aware of crossreactivity!

Labeling of analyte radioactively, fluorescently or enzymatic

Pros:

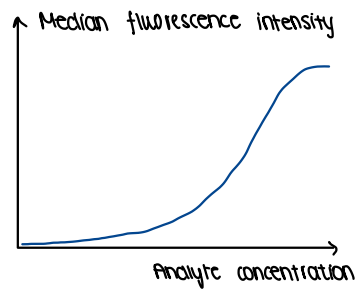
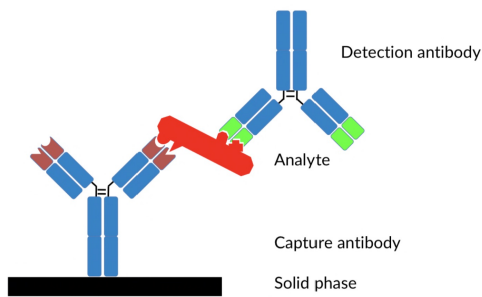
- one antibody
- quantitative
- Suitable for small molecules e.g. pesticides, hormones, ...

Cons:

- radioactive reagents
- specificity
- narrow dynamic range

Non-competitive assay

• Principle Sandwich immunoassay:



→ most prominent example:
SARS COV-2 antigen test

- ↳ Pro's:
- quantitative
 - fast, robust
 - two independent binding events → high specificity
 - signal amplification → high sensitivity

- ↳ Con's:
- two antibodies required

Direct Immunoassay

• Examples: Immunohistochemistry (IHC), serological assays, auto-immundiagnosics

• Principle: antigen is immobilized on a surface & then detected with labelled antibodies

f.e. HER2 (protein in breast cancer) detection in a sample with labeled antibodies → staining of cancer cells

- ↳ Pro's:
- only one antibody
- ↳ Con's:
- labour-intensive
 - subjective - results depend on pathologist

Detection systems

• Direct labeling detection antibody:

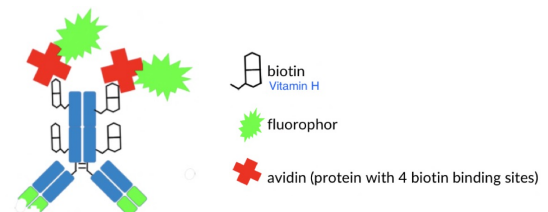
- enzyme e.g. horseradish peroxidase (HRP), alkaline phosphatase
- fluorescent dyes e.g. FITC, TAMRA, polymethine dyes, phycoerythrin

- ↳ one conjugate per assay → no flexibility
- ↳ rapid assay: only two incubation times

• Indirect labeling detection antibody:

- biotin-avidin system

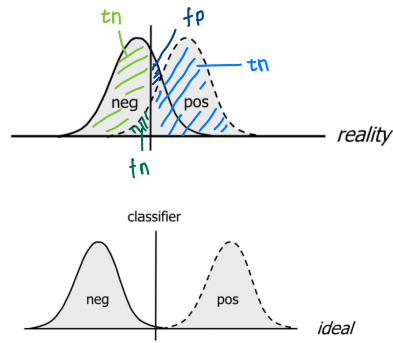
- ↳ one conjugate for many assays → flexibility
- ↳ signal amplification, up to 20 x
- ↳ > 1 conjugate binds to detection antibody
- ↳ capture & detector from one species in one assay but detection antibody needs to be chemically modified



- enzyme-conjugated secondary antibody
- fluorescently labelled secondary antibody
- ↳ one conjugate for many assays → flexibility
- ↳ signal amplification, up to 20 x
- ↳ > 1 secondary antibody binds to detection antibody

Clinical sensitivity and specificity

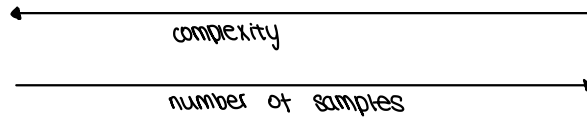
- Sensitivity: $\frac{tp}{tp+fn}$
 tp: true positive
 tn: true negative
 fp: false positive
 fn: false negative
- Specificity: $\frac{tn}{tn+fp}$



Definition biomarker: "A characteristic that can be measured and evaluated as an indicator of normal biologic processes, pathologic processes or pharmacologic responses to therapeutic intervention..."

"... any substance, structure or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease ..."

• Biomarker Identification: Discovery → Qualification → Verification → Validation



Possible exam questions

1. Define the term "epitope" or "paratope"?

Epitope = part of antigen that interacts with the antibody

Paratope = part of antibody that interacts with the antigen

2. Explain the terms "continuous" and "discontinuous epitope"?

continuous epitope = amino acid sequences that are linear in shape

discontinuous epitope = amino acid sequences with a folded conformation

3. Which reagents are required for a sandwich immunoassay?

Solid phase, immobilized primary antibody for capture, analyte, secondary antibody for detection

4. Describe the difference between competitive and non-competitive assay concept?

competitive assay = use of a labeled standard & analyte → inhibition through standard

non-competitive assay = only analyte which is labeled

5. What is detected in a serological assay?

virus or bacteria-specific antibodies are detected

6. What is clinical specificity?

Clinical specificity = ability of a test to correctly identify patients without a disease (true negative rate) → $\frac{tn}{tn+fp}$

7. What is clinical sensitivity?

Clinical sensitivity = ability of a test to correctly identify patients with a disease (true positive rate) → $\frac{tp}{tp+fn}$

7. WESTERN BLOT

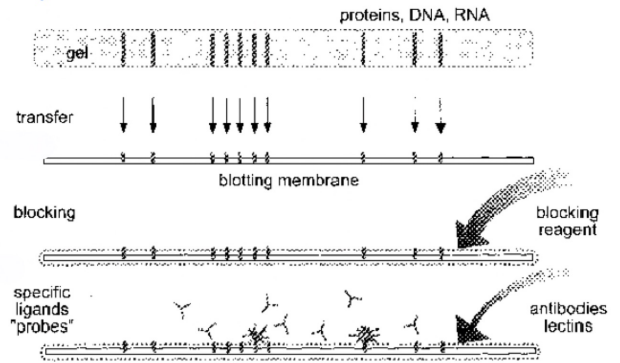
- Prof. Oliver Pötz

09.11.2021

Western Blot concept

= Separation method to detect a protein of interest

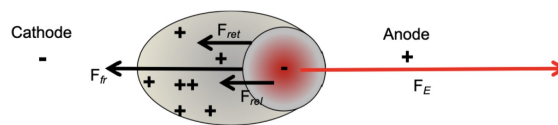
- 4 Steps:
1. Separation (Electrophoresis)
 2. Transfer onto membrane (Blotting)
 3. Blocking
 4. Detection



Electrophoresis

- Definition: - applying a voltage between two electrodes to achieve movement of charged molecules, particles, cells, ... in a constant electric field
 - differentiation of carrier-free (capillary electrophoresis) & gel electrophoresis
 - separation based on differently charged masses & shapes (molecular forms)
- History: - Arne Tiselius received 1948 the Nobel prize in chemistry for his work on electrophoresis & adsorption
- Principle: Particles different in charge & size move in the electrical field with different velocities. Substances move in discrete zones

4 different -forces influence the electrophoretic mobility:



F_E : electrostatic force

F_{fr} : friction force retards ion ($\sim$ velocity, viscosity & pore size, molecule size)

F_{ret} : retardation force (ion cloud-induced solvent polarization retards ion movement)

F_{rel} : relaxation force (ion cloud retards central ion)

• Polyacrylamide gel electrophoresis (PAGE)

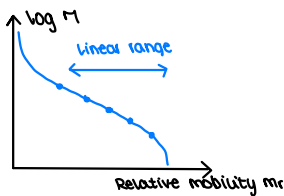
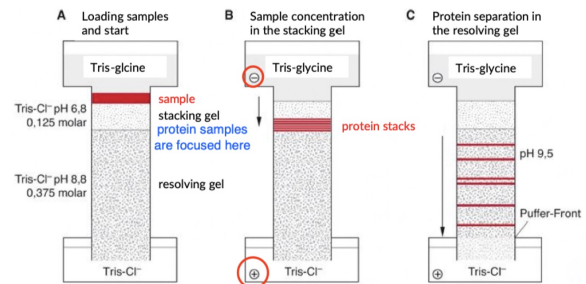
- most common form of gel electrophoresis
- water-soluble synthetic linear polymers made of acrylamide or acrylamide & acrylic acid
- characteristic values:
 - Acrylamide concentration (T)
 - Cross-linker concentration (C) $\rightarrow$ N,N'-Methylenebisacrylamide (Bis)
 - $\hookrightarrow$ ratio between T & C determines the pore size of the polymer, typical mixture 30:1
- Initiator / radical starter (Ammonium persulfate 1-10mM) & catalyst (TEMED 1-10mM) are needed
- characteristics:
 - stable, transparent, nearly no electro-endosmosis
 - excellent separation
 - storable
 - suitable for many staining methods
 - $\hookrightarrow$ monomers are toxic, pore size is limited
- impact of T and C on the separation of a native protein mixture:
 - 47: T no sieving $\rightarrow$ separation based on free mobility (charge, shear coefficient)
 - 87: T sieving effect $\rightarrow$ charge, shape & size-based separation

• SDS-PAGE protein separation based on size

- anionic detergent masks protein charges (normally negative)
 - protein separation according to their molecular mass
 - negative surface charge prevents protein aggregation
 - allows fast separations
- use of 2x Laemmli Buffer (125 mM TRIS-HCl, 4% SDS, 20% glycerol, 0.004% bromophenol blue, 10% 2-mercaptoethanol)
 - ethanol is used to get rid of the surface bonds

• Discontinuous electrophoresis (Disc-electrophoresis)

- sample is load on top of the resolving & stacking gel
- proteins sample is transferred into protein stacks through addition of a voltage
 - ↳ protein is separated in the resolving gel based on ion mobility (small proteins at the bottom, bigger proteins at the top)
- method prevents aggregation & shows higher resolution
- linear range depends on the molecular weight



Relative mobility $\neq$ relative distance of the protein to the starting point
 → Highest resolution in combination with DISC-PAGE

- depending on the selection of T the separation range can be defined

• SDS-PAGE staining methods

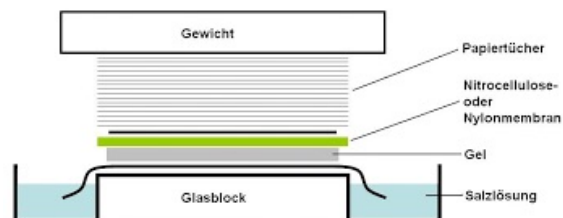
- Coomassie-blue (anionic triphenyl dye): simple, relatively sensitive
- Silver staining: protein binds silver cations which are reduced to elementary silver, silver then forms seed crystals for further chemical redox reactions
 - ↳ very sensitive, requires clean & precise work
- Fluorescence dyes: very sensitive

Blotting

- History: - invention of "Southern blot" by Edwin Southern

↳ method for the detection of specific DNA

1. Partial DNA cleavage by restriction enzymes
2. Agarose gel electrophoresis
3. Blotting onto membrane
4. Hybridisation of complementary labelled probe
5. Removal of unbound probe
6. Detection of bound probe



- Blotting - "atlas": Northern blotting for RNA
 Southern blotting for DNA
 Western blotting for proteins

} all membrane-based methods

• Blotting methods

Methods	Physico-chemical principle
Diffusion blotting	Diffusion (not efficient)
Thermal blotting	Diffusion & temperature (long term, but more efficient)
Capillary blotting	Capillary forces
Pressure blotting	Capillary forces & pressure
Vacuum blotting	Vacuum
Electro-blotting	Electric field
Tank blotting	
Semi-dry blotting	(highly efficient)

→ Size does matter in blotting, i.e. large proteins could only be partially transferred & small proteins could pass the membrane

→ for best sensitivity adjust transfer time according to size of analyte!

Blocking

- Blocking membranes provide high binding capacity → uncoated space has to be blocked
- Examples for blocking reagents: TRIS-buffered saline + 0.05% Tween-20 + 5% skimmed milk powder / 2-10% BSA / 3% gelatin

Detection

- 1st step: specific primary antibody
- 2nd step: visualization of specific primary antibody
 - Non-enzymatic: biotinylated secondary antibody, fluorescently labelled secondary antibody, labelled protein A
 - Enzymatic: enzyme-conjugated secondary antibody → chemoluminescence, fluorescence, dye
 - biotinylated secondary antibody → enzyme-avidin / streptavidin conjugate, dye detection or chemoluminescence
 - enhanced detection by enzymatic signal amplification
- good antibodies: one strong signal at respective molecular weight, high specificity
- medium antibodies: strongest signal at molecular weight, but also other signals
- bad antibodies: second antibody generated with different antigen, validate one antibody with another

Antibody databases

- most common problems with antibodies:
 - cross reactivity
 - variability
 - wrong application
- Information about suitable antibodies can be found on several internet pages

! Possible exam questions

1. Describe the basic concept of gel electrophoresis.
Gel electrophoresis is a method to separate DNA, RNA or proteins depending on their molecular weight by applying a voltage.
2. Describe the basic concept of SDS-PAGE.
SDS-PAGE is a method to separate proteins depending on their size (and not their charge) by using the anionic detergent SDS.
3. How does a stacking gel work?
The stacking gel lines up all the protein samples so that they can enter the resolving gel all at the same time.
4. How can you stain proteins in a SDS-gel?
With Coomassie-blue, with silver cations or with fluorescence dyes.
5. What do you need for Western blot?
A protein sample, gels, a voltage, dye, blocking, primary & secondary antibodies

8. MASS SPECTROMETRY I - THEORY - Dr. Anne Zeck

13.12.2021

Principle of a mass spectrometer

- Mass spectrometer as medical device:
 - identify the exact mass of an analyte
 - structure elucidation of biomolecules (molar mass, fragmentation, isotopic pattern)
 - absolute quantification

advantages & disadvantages:

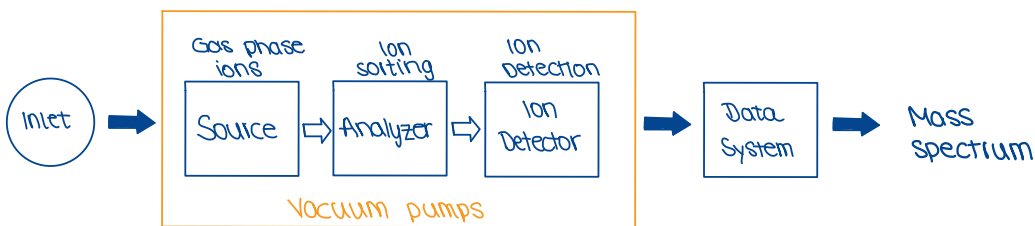
⊕

very highly selectivity
fast & sensitive
low reagent cost
(compared to immunoassays)

⊖

expensive & requires extensive knowledge & experience
only under vacuum possible
salts are not tolerated (e.g. physiological buffers)

Components of a mass spectrometer



a. Ionization of biomolecules

1. Electrospray ionisation (ESI)

- = ions are produced by using an electrospray in which a high voltage is applied to create an aerosol
- ↳ it may produce multiple-charged ions leading to an extension of the mass range of the analyzer
- ↳ accommodation of proteins & their associated polypeptide fragments in a k-mDa range

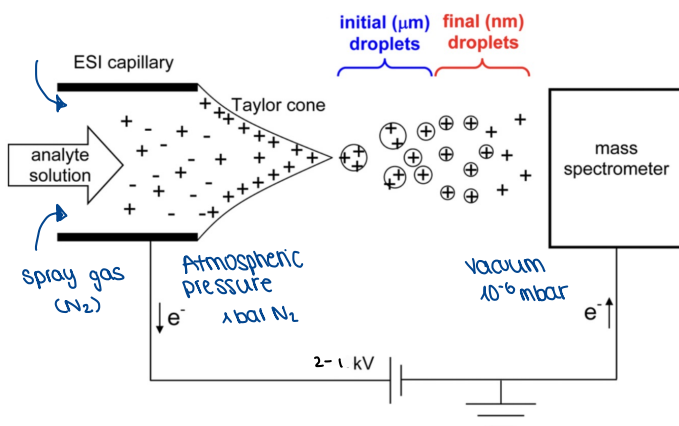
↳ advantages & disadvantages

⊕

- soft ionization under atmospheric pressure by protons
 - ↳ acidic solvents (ACN/H₂O 10.05% formic acid)
- coupling with HPLC & capillary electrophoresis possible
 - ↳ online detection

⊖

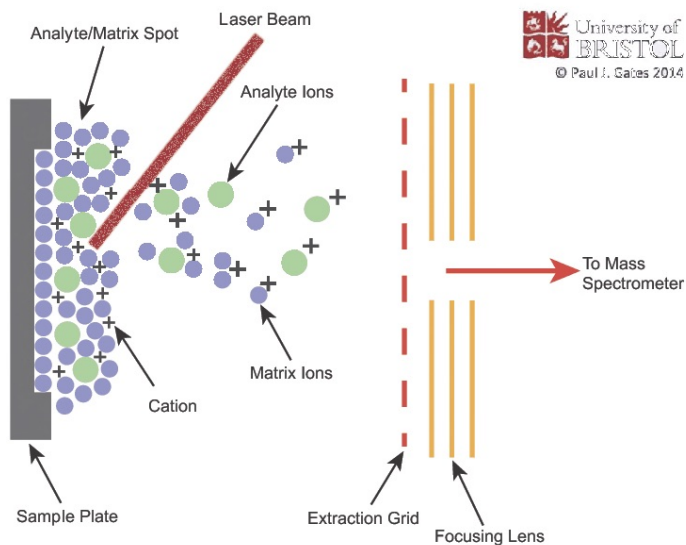
- multiple charges per analyte molecule possible
 - ↳ complex spectra
- only volatile buffer components compatible
 - ↳ limited separation buffer, no ion pairing reagents



- ↳ very thin needles are used, e.g. Nano-spray needle, drawn out glass capillary & „open“ capillary
- ↳ comparison of ionisation efficiency: Electrospray with 25 $\mu\text{L}/\text{min}$ flow rate & 0.01 ion efficacy
Nano-electrospray with <500 nL/min flow rate & 50-60% ion efficacy
↳ low flow rate over small distance is more efficient

11. Matrix-Associated Laser Desorption/Ionization (MALDI)

- = laser energy absorbing matrix is used to create ions from large molecules with minimal fragmentation
- ↳ used for molecules that tend to be fragile
- ↳ MALDI typically produces far fewer multi-charged ions compared to ESI



Mechanism:

1. Formation of a "solid" solution
 - ↳ Matrix is in excess; analyte molecules are isolated from each other
2. Matrix excitation
 - ↳ laser beam excites the chromophore of the matrix
 - ↳ vibrational excitation
 - ↳ desintegration of solid solution spot
 - ↳ matrix evaporates away & leaves free analyte in gas phase
3. Analyte ionisation
 - ↳ cation attachment (Na^+) to analyte by proton transfer

- 4 Required matrix properties:
 - strong absorption of laser wavelength
 - highly sublimable
 - proton donor
- } small organic acid

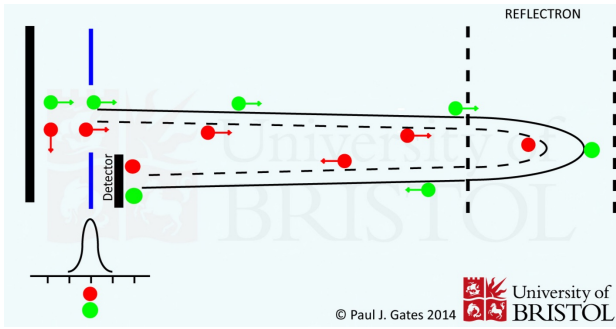
↳ advantages & disadvantages:

- | | |
|--|---|
| <p style="text-align: center;">⊕</p> <ul style="list-style-type: none"> • fast → high throughput • very sensitive detection • extreme less sample consumption • easy to interpret data • relative cheap costs of operation • samples can be reanalyzed | <p style="text-align: center;">⊖</p> <ul style="list-style-type: none"> • not quantitative • no easy online LC coupling • low molecular weight • components not detectable due to matrix background |
|--|---|

b. Separation of the ions

- Examples for analyzer: Quadrupole Time-of-flight (QTOF), Triple Quadrupole or Linear ion trap, Ion Trap, Orbitrap, Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (FT-MS)
↳ based on mass-to-charge ratio

I Time of flight mass analyzer

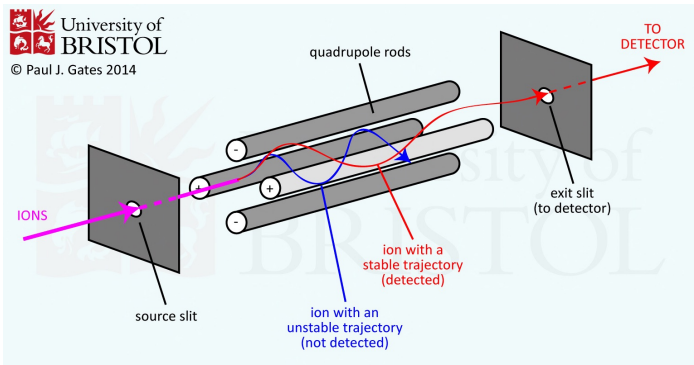


1. Ions are introduced from source (MALDI) or previous analyzer (quadrupole) & clock is started
2. All the ions receive the same initial kinetic energy using an electric pulse
3. Ions drift along the field free drift zone & are separated by their charge / mass ratio
4. Ions reach the detector after different flight times
5. Calculation of m/z from their flight time

E : extraction pulse potential d : length of field free drift zone
 s : length of flight tube t : measured time-of-flight

TOF equation: $\frac{m}{z} = 2eEs\left(\frac{t}{d}\right)^2$

II. Quadrupole mass analyzer



Principle:

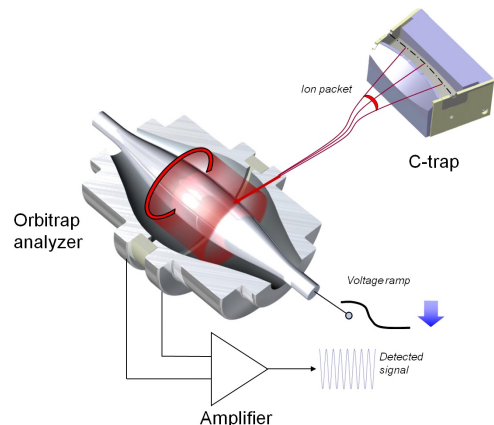
- consists of four parallel rods that have fixed DC & alternating RF potentials applied to them
- Ions produced in the source of the instrument are focussed & passed along the middle of the quadrupoles
- motion will depend on the electric fields so that only ions of particular m/z will have a stable trajectory & thus pass through the detector

• RF is varied to bring ions of different m/z into focus on the detector & build up a mass spectrum

• Pre- & Post-filters are used to isolate stable (resonant) ions → focussing of ion beam

III. Orbitrap mass analyzer

- Principle:
 - ions are transferred from the C-Trap are captured by rapidly increasing the voltage on the center electrode
 - trapped ions assume circular trajectories around the center electrode
 - axial oscillations along the center electrode are detected
- Advantages:
 - very high mass resolution
 - high sensitivity of linear ion trap



⇒ Combination of different analyzers enables to obtain the desired properties for different analysis types

	I. Ion trap	II. Quadrupole	III. Time-of-flight	IV. Orbitrap
Sensitivity	+	+	+	(+)
Speed	+	-	+	(+)
Resolution	-	-	+	+
Mass accuracy	-	-	(+)	+
Mass range	-	-	+	(+)
Fragmentation	+	+	-	(-)

Mass spectrum of peptides

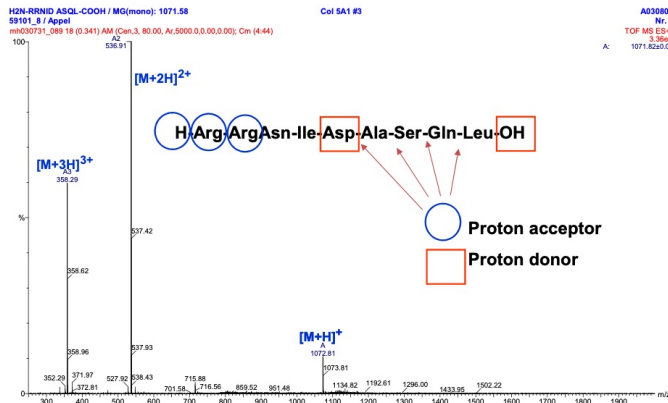
• Example: H-Arg-Arg-Asn-Ile-Asp-Ala-Ser-Gln-Leu-OH

↳ Monoisotopic Mass: 1071.58 Da

↳ Elemental composition: $C_{43}H_{77}N_{17}O_{15}$

↳ Molecular weight: $43 \times 12 + 77 \times 1 + 17 \times 14 + 15 \times 16$

↳ ESI mass spectrum:



mass Spectrometer detects only protonized peptides

↳ Statistical distribution of charges due to statistical attachment of protons

↳ MALDI mass spectrum • MALDI process results in singly charge molecular ions

↳ Spectra easy to interpret

• One peptide = multiple signals in the mass spectrum

• Elemental isotopes: one element shows multiple signals in the mass spectrum

↳ isotopes are separated

↳ calculation of the molecular mass using the different isotopic masses results in multiple masses per sum formula (isotopic distribution)

↳ calculation of the molar mass using the average elemental mass results in a defined mass for each sum formula

• Isotopic distribution:

Monoisotopic mass = mass of a molecule that contains only most abundant isotopes

Average mass = sum of the masses of it's isotopes, each multiplied by it's natural abundance

↳ isotopic distribution for peptides can be used for the identification of the charge state

Singly charged = isotope distance 1 amu

doubly charged = 1/2 amu

triply charged = 1/3 amu

↳ For high molecular weight proteins only the average mass can be determined, the monoisotopic mass is not detectable anymore.

• Example: $\Delta 0.2 \text{ Da} \rightarrow [M+5H]^{5+}$

Average mass = 1567.4685 (5+)

Monoisotopic mass = 1566.5274 (5+)

Res: 0.1 amu

↳ mass peptide without protons: $\frac{m_{mi}}{Z} \cdot Z - Z \cdot m_{proton}$

$$\frac{1566.5274 \text{ Da}}{5} \cdot 5 - 5 \cdot 1 \text{ Da} = 1561.5274 \text{ Da}$$

9. MASS SPECTROMETRY II - APPLICATIONS

- Dr. Anne Zeck

20.12.2021

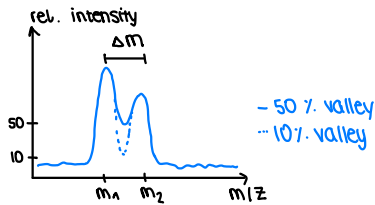
Mass spectrometry of proteins

- Mass resolution = capability of a mass spectrometer to resolve two ions with minimal mass difference

$$R = \frac{m}{\Delta m} = \frac{m_1}{m_2 - m_1}$$

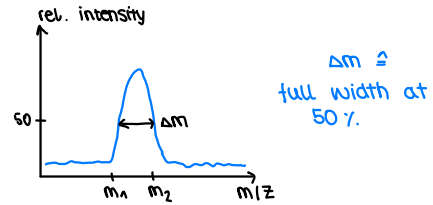
↳ different definitions for different mass spectrometers:

A) Quadrupole & sector field mass spec.



R defined by peakpair

B) Time-of-flight (TOF)



R defined by one peak

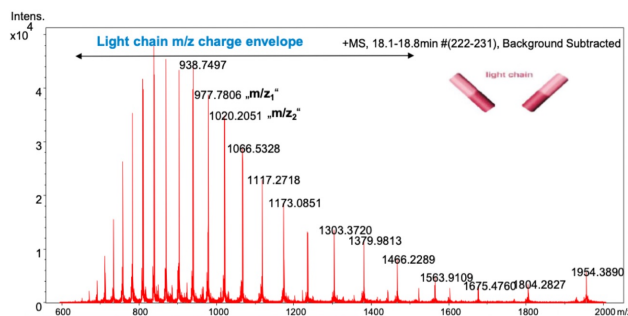
↳ The resolution of the isotopic pattern increases with increasing mass resolution

- Mass resolution for protein peaks: The larger the biomolecules are, the more charges they retain during electrospray process & the closer the isotopic m/z peaks become
 - The larger the protein, the higher needs to be the mass resolution in order to calculate the charge in this way
 - At ~ MW of 50,000 Da (e.g. heavy chain of an antibody) it is not possible to resolve singly isotopic peaks any more

Example: ESI spectrum of the light chain of an antibody (25 kDa)

- charge deconvolution $z = \frac{(m/z)_2 + 1}{(m/z)_2 - (m/z)_1}$
- average mass $MW = \frac{(m/z)_2 + z}{z}$

↳ calculation for two neighbouring peaks
 $(m/z)_1 = 977.78$ Da, $(m/z)_2 = 1020.20$ Da,
 $MW = 23442.74$ Da



$$\frac{m}{z_1} = \frac{\text{protein molecular weight} + z \cdot 1}{z} = 977.78 \text{ Da}$$

proton molecular weight

protein charge state

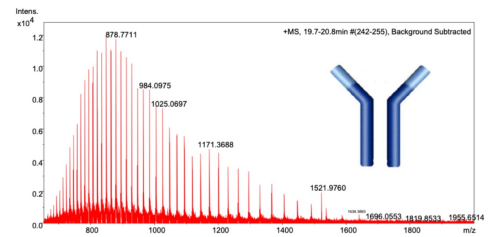
$$\frac{m}{z_2} = \frac{MW + (z-1) \cdot 1}{z-1} = 1020.20 \text{ Da}$$

adjacent peak with one charge less on the protein

- ↳ charge envelope: intensity reflects statistical probability of the uptake of one positively charged proton during the electrospray process
- ↳ isotopical resolution of antibody light chain with FT-ICR, Orbitrap & modern Q-TOF instruments → monoisotopic mass can be determined
- ↳ Deamidation ($\Delta m = 1$ Da) can be detected at the level of light chain

• Example: ESI spectrum heavy chain of an antibody (50 kDa)

- Signal density shows a highly charged protein (~50)
 - ↳ lower space between the peaks compared to light chain
 - ↳ number of peaks HC = 2 × number for LC



- charge deconvoluted heavy chain does not show isotopic peaks anymore
 - resolution too low → average mass is used for large proteins

• Detection of protein degradation: Mass spectrometry allows (besides the confirmation of protein identity) the elucidation of degradation pathways & the quantification of posttranslational modifications

e.g. Oxidation, non-enzymatic glycation

• Determination of glycosylation: Glycans may have an impact on solubility, immunogenicity or pharmacokinetics of a protein

↳ mass changes can be detected

Mass spectrometry of peptides

• Peptide mass fingerprinting = analytical technique for protein identification in which the unknown protein is cleaved into smaller peptides, whose absolute mass can be measured with mass spectrometry

↳ Procedure: 1) Enzymatic digestion of the protein

2) Chromatographic separation of the peptides & mass measurement by MS

↳ HPLC-ESI-MS analysis enables the identification of peptides by retention time & molecular mass

↳ EIC = extracted ion chromatogram → only one m/z is displayed

TIC = total ion chromatogram → all m/z signals recorded at a given time are summed up

3) Identification of the protein in a database by matching masses of a subset of the tryptic peptides against calculated values

• Mass spectrometry of peptide mixtures: If sample is too complex data interpretation of ESI spectra is not possible

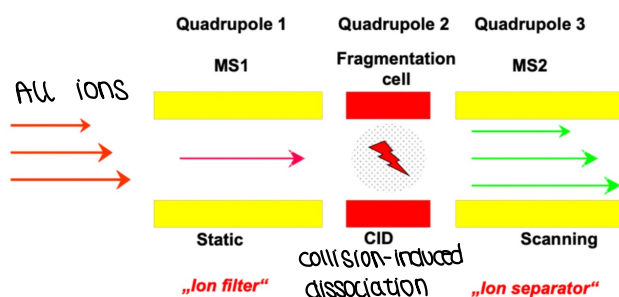
↳ New possibility: fragmentation of the ions within the mass spectrometer to acquire typical fragment ions corresponding to the mass of the amino acid

Mass spectrometry of small molecules

• Tandem mass spectrometer with MS-MS capability



↳ Fragmentation principle: Triple quadrupole instruments

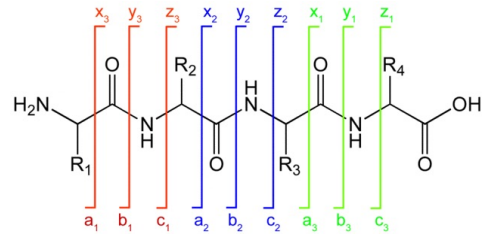


→ Quadrupole 1 & 3 act as mass filters

→ Quadrupole 2 causes fragmentation of the analyte through interaction with a collision gas

Peptide fragment notation:

- ions indicated by a, b, c if charge is retained to the N-terminus
- x, y, z if charge is retained to C-terminus
- subscript i indicates number of amino acid residues in the fragment



4 Attention: The fragmentation of multiply charged ions produces fragments with higher m/z ratio than the precursor ion showed.

4 The difference of fragment masses is characteristic for amino acids & therefore enables to identify them. BUT the terminal amino acid information is usually missing.

Applications of MS in the clinics

• MALDI Biotyper: Mass Spectrometry based identification of microorganisms using their molecular fingerprint

- 4 Procedure:
- 1) Applying a small amount of sample directly on a MALDI target plate
 - 2) Thin microbial film is overlaid with 2 µL of a saturated acid matrix solution
 - 3) Mass Spectra between 2,000 and 10,000 Da
 - 4) Identification of bacteria, yeast, fungi by using the proteomics fingerprint & peak-matching / multivariate analysis.

- 4 Advantages:
- no initial assessment like gram staining, oxidase test of unknown samples or choice of PCR primers
 - 4 faster than PCR or metabolic pathway analysis
 - generic experimental setup
 - robust method, minimal sample preparation

- 4 Applications:
- clinical routine monitoring
 - environmental & taxonomical research
 - food processing & quality control

- Medical devices:
- subdivided into three classes depending on patient risk: class I-III
 - each risk class has its own standard of evidence & requirements for review
 - class I devices used in healthcare applications

4 General requirements for the approval of medical devices:

- 1) Intended use / indications for use
- 2) Analytical performance testing
- 3) Clinical performance testing
- 4) Software
- 5) Device labeling (package insert / instructions for use)

• Summary: MS in clinics

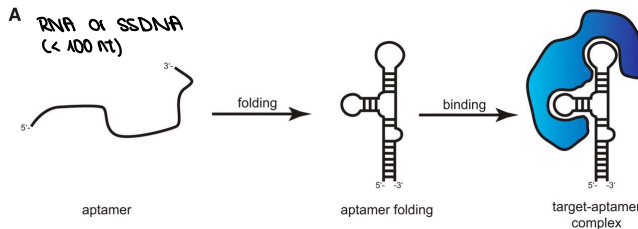
- MS can measure multiple analytes with high sensitivity & selectivity
- MS is often faster & cheaper than other methods
- MS can be classified as "Class I Medical Device" intended for the use in diagnosis

- Applications:
- Endocrines & Biomarkers: Vitamin D, Steroids, peptides
 - Therapeutic drug monitoring: immunosuppressants, pain relief drugs, ...
 - Inborn errors of metabolism: acyl carnitines, amino acids, ...
 - Toxicology research: drugs of abuse, opiates, ...

RNA / DNA Aptamers

- Definition:
 - lat. aptus = to fit, gr. meros = meaning piece → "small fitting pieces"
 - ↳ aptamer = short-stranded DNA or RNA oligonucleotides (~25-90 nucleotides)
 - can be selected out of a huge oligonucleotide database with $\sim 10^{15}$ different sequences

Working principle:

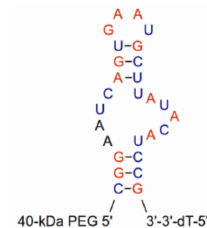


Dissociation constants (K_D) of aptamer-target complexes in pico- to nanomolar range
 ↳ comparable or better than monoclonal antibodies

- Aptamers can be selected against a variety of targets:
 - small molecules
 - peptides & proteins
 - bacteria
 - antibiotics
 - viruses
 - cells
- SELEX = Systematic Evolution of Ligands by Exponential Enrichment

= combinatorial chemistry process to select aptamers

- established 1990 independently by Ellington & Sotinak, Tuerk & Gold, and Robertson & Joyce
- first representative of an aptamer drug: Macugen (Pegaptanib)
 - ↳ highly modified anti-VEGF RNA aptamer for treatment of age-related macular degeneration (AMD)
 - ↳ approved in 2004 by FDA



→ only 2 unmodified ribonucleotides (black)

Aptamer based assay to detect cocaine

- Fluorescence resonance energy transfer (FRET) -based assay
- uses modified cocaine aptamer & quencher oligonucleotide
- working principle:
 - 1) Absence of cocaine
 - quencher oligomer hybridizes to the partially unbound aptamer
 - fluorophore & quencher come into close proximity & fluorescence is detected
 - 2) Presence of cocaine
 - aptamer folds in stable cocaine binding structure
 - quencher oligomer is displaced from the aptamer & released into the surrounding solution
 - fluorescence intensity increases in dependence of the cocaine concentration

Cell-specific aptamers

- aptamers obtained against whole cells:
 - Immunoglobulin heavy μ chain
 - Tyrosine-Protein Kinase-like 7 Enzyme
 - Tenascin C
 - human epidermal growth factor receptor 2 (HER-2)
 - Transforming growth factor-beta type III receptor (TGFRIII)

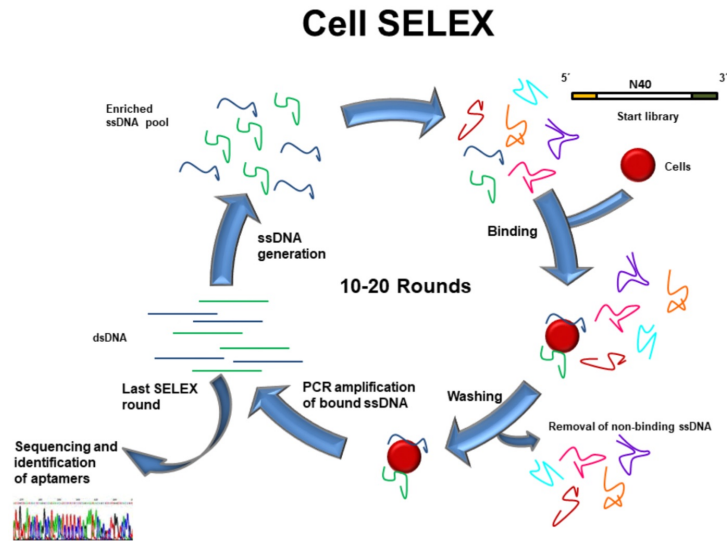
• Cell-SELEX process for selecting DNA aptamers

1. Negative selection

- non-target cells incubated with the combinatorial ssDNA start library consisting of a centrally randomized region of i.e. 40 nucleotides flanked by two specific primer regions
- non-target cells with bound oligonucleotides are removed

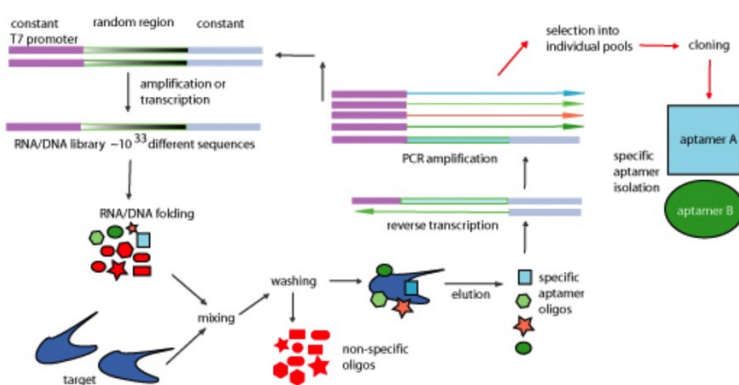
2. Positive Selection

- = remaining ssDNA pool is incubated with target cells
- unbound oligonucleotides are removed during washing steps
- ssDNA bound to the cells is amplified by PCR using fluorescein & phosphate modified primers
- PCR product is purified & ssDNA is generated by lambda exonuclease digestion of phosphorylated undesired strand
- resulting enriched ssDNA pool is used in the next SELEX round for the incubation with cells
- enrichment of target specific aptamers is monitored by flow cytometry
- in last SELEX round, next generation sequencing (NGS) is performed for aptamer identification



• RNA SELEX process

SELEX Protocol



- DNA aptamers more stable than RNA aptamers

• Advantages of Cell-SELEX compared to protein-SELEX

- cell surface proteins are presented in their natural surrounding & native conformation
- cell surface proteins contain possible posttranslational modifications

• Advantages of aptamers

- can be easily & chemically synthesized with high reproducibility & accuracy
- stable in long-time storage & reversible denaturation
- in-vitro selections without prior knowledge of target structure
- target can be identified after selection of aptamers
 - 4 new previously undiscovered marker can be identified
- negative selection with non-target cells possible to remove unspecific binders & to obtain discrete specificities

- precise adaption during their in-vitro selection phase (physicochemical properties)
- aptamer stability & half-life can be increased by chemical modifications of nucleotides at 2' position by using fluoro/amino-modification or coupling to PEG
- reporter molecules can be site-specifically inserted at defined sites
- small size facilitates in-vivo delivery by promoting tissue penetration

• Applications of cell-specific aptamers

- sensors
- diagnostics
- therapeutics: inhibition, stimulation & delivery of cargoes
- imaging
- regenerative medicine

Synthetic modified mRNA

- comparison virus vector: possibility to insert genomic DNA in cells → mutation
- mRNA = single-stranded molecule of RNA that corresponds to the genetic sequence of a gene & is read by the ribosome in the process of protein synthesis
 - ↳ synthetic mRNA can be inserted into the cytoplasm to induce expression of target protein
- Structure of modified synthetic mRNA



5' Cap: against exonucleosis
 5' & 3' UTR: untranslated regions
 Poly-A-tail: for stability

ORF: open reading frame
 = coding sequence (cod.)

• Synthesis of modified synthetic mRNA



- 1) Plasmid insert (CDS of desired protein) is amplified by PCR & a poly-A-tail (110 nucleotides) is added
- 2) PCR & purification of PCR product
- 3) in vitro transcription (IVT) of DNA into mRNA using a T7 RNA polymerase
 - ↳ anti-reverse cap analog (ARCA), ATP, GTP, 5-Methyl-CTP, pseudouridine-5'-triphosphate (ψ) used
- 4) DNase treatment for removal of template DNA
- 5) Antarctic phosphatase treatment to remove 5'-triphosphates
 - ↳ mRNA modification to reduce immune activation
- 6) Quality analysis by agarose gel electrophoresis

• Transfection of cells with modified synthetic mRNA

- transport particles needed because cell surface & DNA are negatively charged
- working principle:
 - 1) Generation of lipoplexes for mRNA transfection
 - 2) mRNA transfected cell
- examples:
 - eGFP transfection → transfected cells show green fluorescence
 - mRNA treatment to express surface receptor



• Generation of induced pluripotent stem cells (iPSCs)

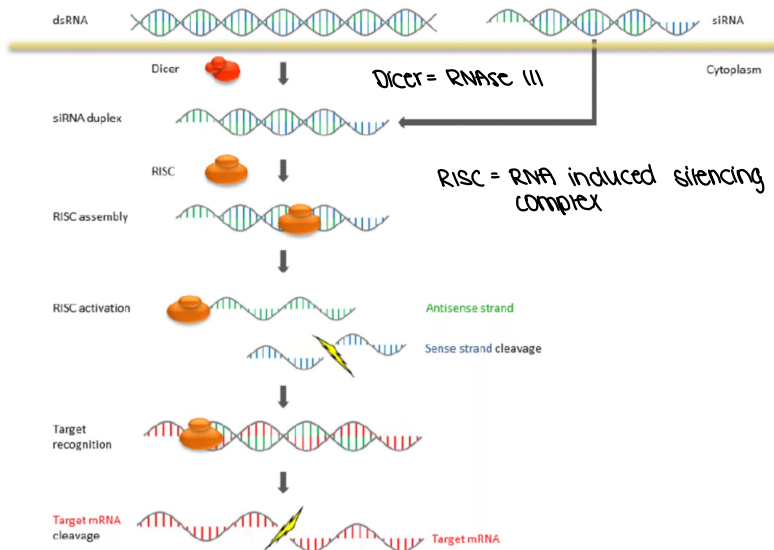
- treatment of fibroblasts with Yamanaka factors (OCT4, KLF4, cMYC, LIN28 & SOX2) coding mRNA
- first colonies visible after 13 days → characterization by fluorescence microscopy, flow cytometry & RT-PCR
- iPSCs form embryoid bodies
 - ↳ Generation of desired autologous cells possible, e.g. beating cardiomyocytes
 - ↳ Tissue regeneration without rejection reaction

Small interfering RNA (siRNA)

- **Definition:**
 - double stranded DNA (dsDNA) with phosphorylated 5' ends & hydroxylated 3' ends with two overhanging nucleotides
 - 21-25 nt long, mostly exogenous origin
 - ↳ plays an important role in the RNA Interference (RNAi) pathway



RNAi pathway



- 1) long dsDNA cleaved by Dicer
- 2) siRNA incorporated into multinuclease complex RISC
- 3) RNA helicase within the complex unwinds siRNA
↳ RISC activation
- 4) cleavage of siRNA sense strand
↳ remaining anti-sense strand - RISC - complex binds to target mRNA
↳ mRNA degradation
- 5) translation & expression of target mRNA are stopped

siRNA applications

- Gene knockdown:
- to study the function of a specific gene
 - for therapeutic gene modulation
 - for antiviral treatment
 - for cancer treatment by tumor cell gene silencing

Challenges: avoiding non-specific effects

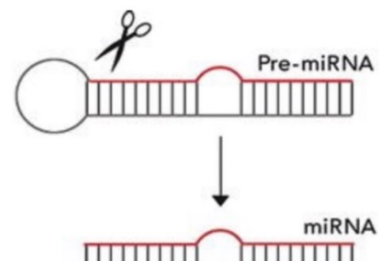
- off-target effects: expression of multiple genes can be reduced
 - ↳ software tools to automatically check cross-reactions
 - ↳ negative control siRNAs (scrambled siRNA)

Targeted delivery of siRNA

- 1) Non-covalent aptamer-siRNA conjugates
- 2) Covalent aptamer-siRNA or -shRNA chimeras:
 - aptamer binds to prostate-specific membrane protein (PSMA)
 - after endosomal release, chimeras are cleaved by Dicer
 - ↳ siRNAs enter RNA-induced silencing complex (RISC) pathway
- 3) Covalent aptamer-nanoparticle conjugates

MicroRNA (miRNA)

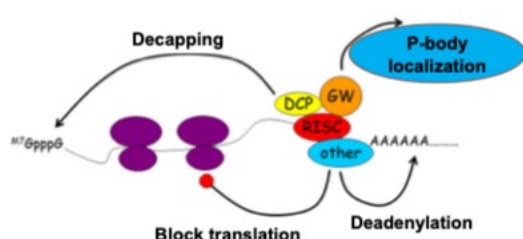
- **Definition:**
 - short, single stranded RNA molecules (~22 nt)
 - mostly endogenous
 - ↳ miRNAs regulate the expression of genes affecting multiple cellular functions including proliferation & growth



• Transcription & processing

- miRNA genes transcribed by RNA polymerase II as large primary miRNA (pri-miRNA)
- pri-miRNA processed by a protein complex containing the RNase III enzyme Drosha, the dsRNA binding protein DGCR8
 - ↳ formation of precursor miRNA (pre-miRNA), ~70 nt
- Pre-miRNA transported in cytosol & processed by Dicer
 - ↳ generation of ds miRNA duplex, ~22 nt
- Dicer is rapidly unwound as it associates with Argonaute (Ago) proteins
 - ↳ one strand is retained to become mature miRNA & loaded into RISC
- mRNA can either be translationally repressed (partial complementarity) or cleaved (perfect complementarity)
 - ↳ depending on level of complementarity between miRNA & target sequence

• Mechanisms for translational repression



Processing bodies (P-bodies) = cytoplasmic granules containing protein-RNA complexes

- ↳ in P-bodies, miRNA targeted mRNA can be sequestered from the translational machinery & degraded or stored for subsequent use

• miRNA & gene expression

- individual miRNAs target ~100 different mRNAs due to partial base complementarity
- individual mRNAs may contain multiple binding sites for different miRNAs
 - ↳ complex regulatory network

• Function of miRNAs

- cell cycle control
- apoptosis
- stem cell differentiation
- immune responses
- hematopoiesis
- muscle development
- neurogenesis
- insulin secretion
- cholesterol metabolism
- aging

→ highly tissue specific expression & distinct temporal expression patterns during embryogenesis
↳ differentiation & maintenance of tissue identity

→ highly involved in diseases, e.g. cancer, heart disease & neurological diseases

• Application of miRNAs

- to restore a loss of function (miRNA replacement)
- miRNA antagonists (anti miRNA) to inhibit endogenous miRNA in diseased tissue
- diagnostic & prognostic biomarkers for diseases

↳ Vascular medical engineering:

- DNA aptamers to capture molecules for in vivo endothelialization
- siRNA / mRNA for gene silencing or induction of protein synthesis

Exam questions

1. How can DNA aptamers be selected against cells (SELEX)?
2. How can modified mRNAs be synthesized?
3. Structure & function of siRNA?
4. What is RNAi?

11. Detection & Sensor Principles

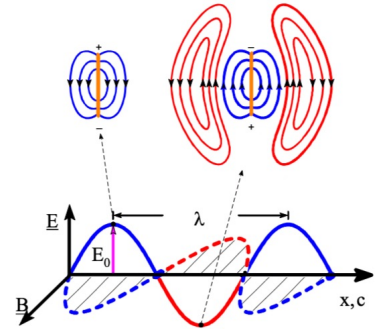
- Prof. G. Gauglitz

17.01.2022

Radiation

Electromagnetic radiation:

- Electric & magnetic field perpendicular to each other & the propagation direction
- At origin of radiation dipole radiation has a phase difference of $\pi/2$ → „near field“
- Field vectors are in phase (vacuum)
- Magnetic field propagates as a ring perpendicular to direction of dipole
- de Broglie: Dualism of particle & wave



Properties of electromagnetic radiation: Amplitude, Frequency, Phase, State of polarization, Time dependence

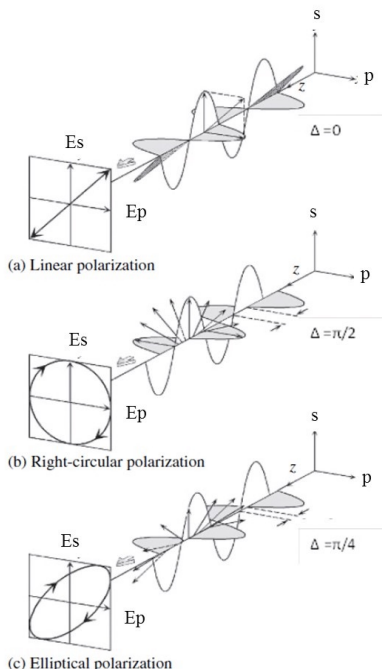
↳ Possible Interactions: Absorption / Transmission, Emission (Fluorescence, Phosphorescence), Scattering (Rayleigh, Raman), Reflection, Refraction

Optical Spectroscopy:

- UV: 200 - 380 nm
- visible: 380 - 780 nm
- NIR: 780 - 1400 nm
- IR: 1400 - 3000 nm

Polarisation = direction of oscillations

- linear: $\Delta\varphi = 0$, $|E_x| = |E_y|$
- circular: $\Delta\varphi = \pi/2$, $|E_x| = |E_y|$
- elliptical: $\Delta\varphi \neq 0$, $|E_x| \neq |E_y|$



Optical rotation - dispersion & radiation

- specific rotation $[\alpha_\lambda] = \text{ORD} = \frac{\alpha_\lambda}{dc}$
- Cotton-Effekt $\theta = \frac{2,303}{4} \cdot \frac{180}{\pi} (\epsilon_e - \epsilon_r) cd$
- Ellipsometry & water control: refraction

Interference

- constructive
- destructive
- beat waves

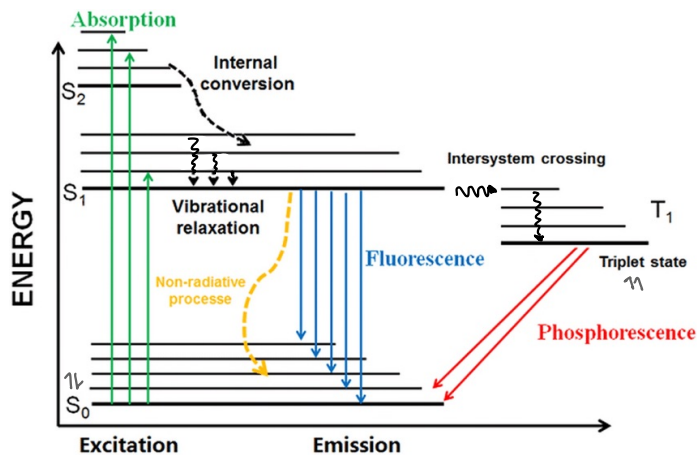
Coherence in time & space

- time: e.g. beat frequency
- space: e.g. two non-correlated waves result in average intensity of zero

Interaction: radiation - matter

- Interactions:
 - Absorption / Transmission: $I_\lambda(d) = I_\lambda(0) \cdot \exp(-d \cdot \epsilon_\lambda \cdot c)$
 - Fluorescence: $I_\lambda^F(t) = I_\lambda(x) \cdot \epsilon_\lambda \cdot c(t) \cdot f_R \cdot \eta \cdot \Gamma$
 - Reflection: Lambert cosine law
 - Scattering: Rayleigh (particle size $< \lambda/10$), Mie (particle size $\sim \lambda/10$)

• Jablonski diagram = illustrates the electronic states of a molecule & transitions



- states arranged vertically by energy
 & grouped horizontally by spin multiplicity

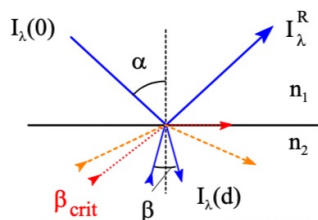
~ = non-radiative transitions
 ↘ = radiative transitions

- i) vibrational relaxation (te)
- ii) internal conversion (ic)
- iii) intersystem crossing (isc)

- vibrational ground states = thick lines
 higher vibrational states = thin lines
 ↳ Relaxation of excited state to its lowest vibrational level = vibrational relaxation

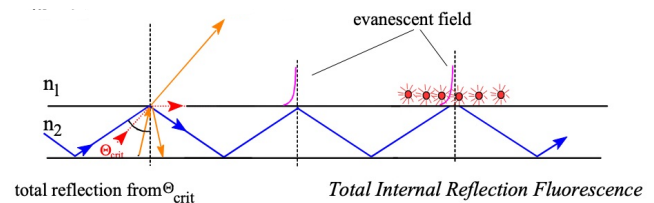
- Absorption: • increase in energy to higher vibrational states
 • no change in spin direction but electrons in different orbitals $S_0\ 1v \rightarrow S_1\ 1v$
 ↳ vibrational relaxation to ground level of S_1 & then non-radiative transition
 - Fluorescence: • decrease in energy from S_1 to S_0
 • shorter wavenumbers in spectrum
 - Phosphorescence: • decrease in energy from T_1 to S_0
 • shorter wavenumbers & intensity in spectrum
- ↳ change of spin is not allowed for radiation, therefore e.g. $S_0 \rightarrow T_1$ transition (normally) not possible

• Reflectance: $-n_1 < n_2$ & $\alpha > \beta$:



$n_{low} \rightarrow n_{high}$: reflection
 $n_{high} \rightarrow n_{low}$: refraction

- $n_2 > n_1$:

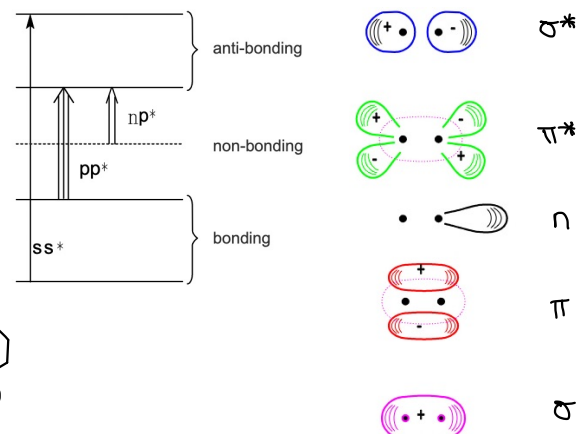


total reflection
 β_{crit}

evanescent field interactions
 e.g. fluorescence in surface molecules

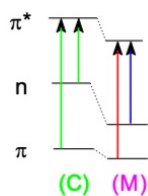
• Orbitals and transitions

- Atomic & molecular orbitals describe the location & wave-like behaviour of an electron
- Three types of orbitals:
 - i) Bonding orbitals
 - ii) Anti-bonding orbitals
 - iii) Non-bonding orbitals



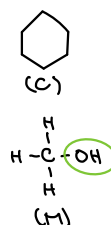
• Band shift in dependence on solvent

- orbital transitions change when solvent is changed from cyclohexan (C) to methanol (M)



↳ energy for orbital transitions change & therefore enable to better distinguish $\pi\pi^*$ & $n\pi^*$ transitions

↳ determination if H atom is in solvent

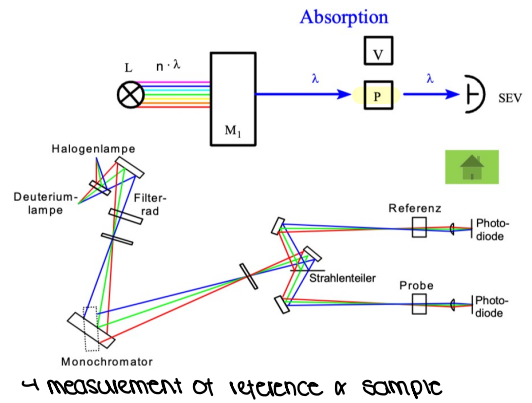


Instrumentation

Absorption

Different classical spectrometer systems using only one specific wavelength, such as UVIKON 8101820

- Nowadays switch to diode arrays that use the whole white light beam. The detector can then detect the different signals of all different colors of the light beam



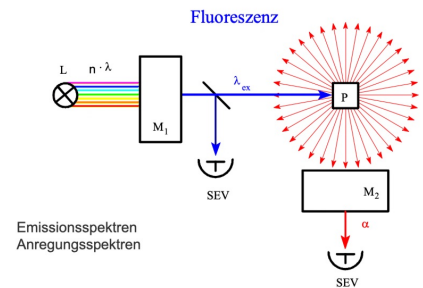
Fluorescence

- Fluorescence is emitted perpendicular over the whole tube around the sample.

- Two different spectrometry measurement methods:

- 1) Select one excitation wavelength in absorption maximum
→ detection of whole fluorescence spectrum
- 2) Select best intensity of fluorescence & change excitation wavelength accordingly
→ detection of excitation spectrum indirectly

- No reference samples: stabilization of excitation intensity with second diode



- Light sources: typically combination of hydrogen & deuterium

- not too good for red-yellow light, therefore use of Tungsten & halogen sources
- for UV & visible: xenon light source
- lasers or LEDs

- Separation of light wavelengths: by a filter or by refraction

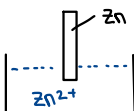
- Prism monochromator: refraction dependent on wavelength → non-linear dispersion
- Grating monochromator: interference effects of refracted light → linear dispersion
→ cut-off filters needed because wavelengths can superpose
→ higher orders = higher resolution

Detectors

- CCD = charged coupled devices
- CMOS = complementary metal oxide semiconductor, e.g. in smartphones
- Diode array produces voltage:
 - linearly proportional to intensity of incident light
 - not linearly proportional to wavelength
- Multi-layer system of silicon:
 - silicon sensitive to light
 - layer structure allows laterally varying functionality

Basics in Electrochemistry

Half-cell:



- Phase equilibrium → Galvani-potential

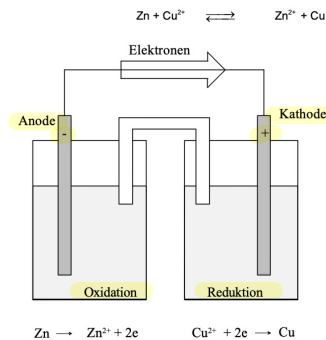
- equilibrium in chemical potential: $\tilde{\mu}_i^\alpha = \tilde{\mu}_i^\beta$

- Galvani-voltage: difference of potentials $\Delta \varphi_0 = \varphi^M - \varphi^L$

• Electrochemical potential

- chemical potential $\mu_i = \mu_i^0 + RT \ln a_i$ & additional Galvani-potential φ result in electrochemical potential $\tilde{\mu}_i^{\alpha} = \mu_i^{\alpha 0} + RT \ln a_i + z_i F \varphi^{\alpha}$
chemic thermodynamic electric

- Daniell-Element:



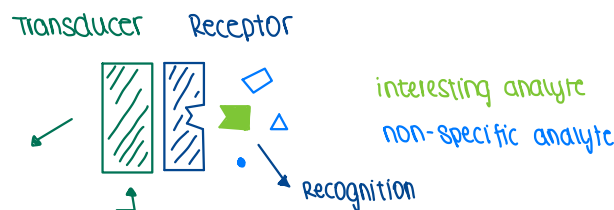
- Standard-Reduction-Potential table (Stockholm-convention) gives data values of standard electrode potentials (E^0) in V relative to the standard hydrogen electrode

$$\Delta E = E^0_{\text{kathode}} - E^0_{\text{anode}}$$

- measurement by digital voltmeter (Poggendorf compensation)

Sensors

- Sensors can be based on temperature, pressure, evoked potentials, biomolecular interactions
- Definition: device that receives & responds to a signal or stimulus, e.g. thermometer, gas detection
- Sensor components: Physical sensor (temperature, pressure, photoelectric barrier)
 & Chemical / biochemical sensor (ion-sensitive electrode, optrode)



- Classification of sensors:

i) Electrochemical ii) Optical iii) Mechanical iv) Thermal	}	General requirements: selectivity, accuracy, repeatability, reproducibility, linear scale, respond quickly, reversibility, small & cheap
---	---	---

- Historic approach:
 - 1953 Brittain & Bardeen first reported gas sensitive effects on Germanium
 - 1954 Heiland found gas sensing effect on metal oxides
 - Taguchi finally brought semiconductor sensors based on metal oxides to an industrial product

• Gas sensors:

- measurement of:
 - electromotive force
 - diffusion & mass transfer with amperometric gas sensors (AGS)
 - potential difference between working & reference electrode with potentiometric sensors

• Thickness-Shear-Mode (TSM) resonators

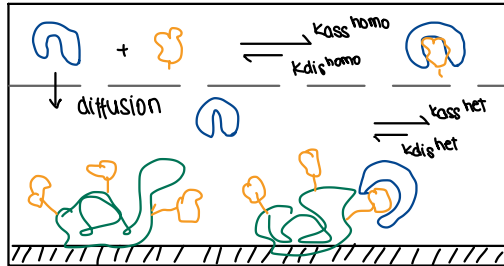
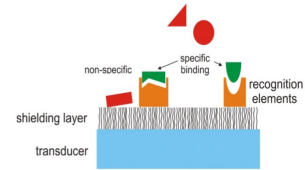
- quartz sensors that show a frequency shift due to increased mass
- based on piezoelectric effect → electrical signal converted into an acoustic wave

12. Biosensors, Assays and BIA - Prof. G. Gauglitz

24.01.2012

Biosensors

- **Biosensor principles**:
 - i) Physical sensor → temperature, pressure, light barrier
 - ii) Chemical / Biochemical sensor → ion sensitive electrode, optode
- **Biomolecular interaction**: Different kinetics in homogeneous & heterogeneous phase



homogeneous phase

diffusion layer

heterogeneous phase
surface modification
transducer

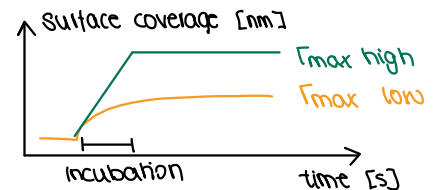
U = recognition element

Y = analyte

- **Time resolved & label-free measurements**:
 - Qualitative measurement → binding / no binding
 - Quantitative measurement → concentration
 - Kinetics → association & dissociation rate constants
 - Thermodynamics → affinity & binding constants

↳ Detection of interaction kinetics dependent on surface coverage

- Γ_{max} high = high loading → diffusion
- Γ_{max} low = low loading → kinetics at binding surface

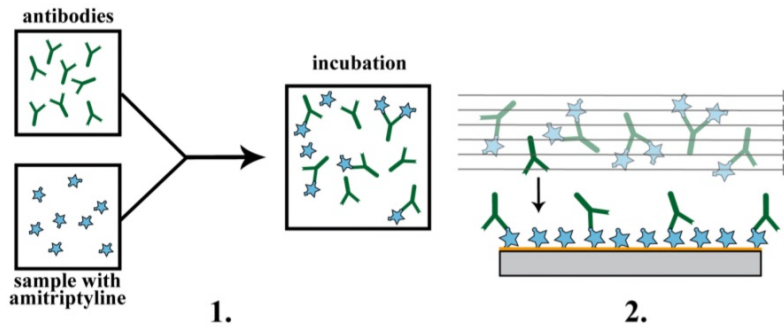


Assays

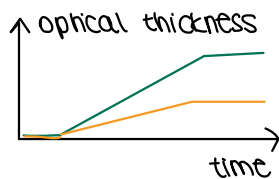
- **Homogeneous phase**:
 - i) Scattering → Binding of recognition element & analyte leads to change of scattering behaviour
 - ii) Quenching → Fluorophore fixed on recognition element will be quenched during binding process with analyte
 - iii) Fluorescence Resonance Energy Transfer (FRET)
 - Different fluorophores fixed on recognition element & analyte. Binding leads to energy transfer by dipole interactions & therefore switches fluorescence wavelength
- **Competitive assay**: competition of two different binding partners
→ signal only for one binding partner received
- **Binding inhibition assay**: analyte fixed on surface, but also in solution
 - recognition element binds to analyte in solution = no signal
 - recognition element binds to immobilized analyte = signal
 - ↳ big advantage: analyte can be very small
- **Sandwich assay**: primary recognition element bound to surface, secondary recognition element binds to already attached recognition element-analyte complex
 - measurement of increased surface mass or fluorescence possible
 - ↳ but: only for big analytes as two binding sites are required

- **Direct assay:** recognition element immobilized at surface
 - binding of analyte = signal
 - simple for large analytes

• Flow injection analysis of binding inhibition assay



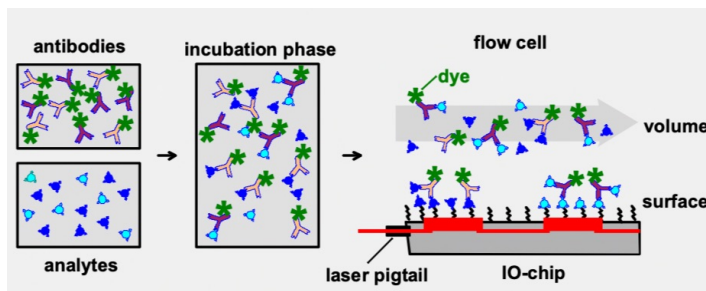
- 1.) Incubation of antigen with defined amount of antibodies
 - ↓ concentration = ↓ blocked A&S
 - ↑ concentration = ↑ blocked A&S
 - 2.) Incubated solution is put on transducer surface
- ↳ inverse information signal



high increase = ↑ signal = ↓ analyte concentration
 low increase = ↓ signal = ↑ analyte concentration

- ↳ Summary:
- Direct test format = heterogeneous (diffusion)
 - Binding inhibition test either labeled or label-free = homogeneous or heterogeneous

• Antigen-Antibody interaction



Detection based on TIRF
 (Total internal reflection fluorescence)

- ↳ inverse information:
- ↑ analyte con. = ↓ signal
 - ↓ analyte con. = ↑ signal



Recognition elements and surface chemistry

- **Biomolecular interactions by:**
 - electrostatic, hydrophobic or Van der Waals interactions
 - highly ordered arrangement of parameters

↳ Example: DNA immobilisation & on-line monitoring of hybridisation

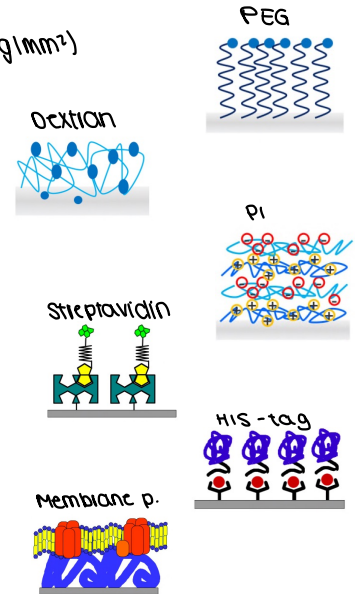
→ Reduction of affinity with increasing number of non-matching bases

- **Optimum surface:**
 - supplies:
 - possibility of high loading
 - avoiding non-specific binding
 - high stability
 - surface modification
 - multi-layer
 - localised spots
 - multiple regeneration

- **Surface chemistry:** providing functional groups for direct immobilisation of recognition elements for further functionalised layers, e.g. biopolymers

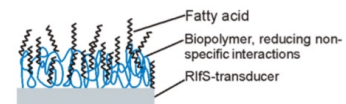
→ Examples for surface modification:

- Silanes
- Polyethyleneglycol:
 - low binding capacity, max. 1 binding layer (5-6 ng/mm²)
 - NH₂ or COOH functionalised polymer
- Dextran:
 - flexible hydrogel
 - high binding capacity (10 ng/mm²)
 - NH₂ or COOH functionalised polymer
- Polyionic multilayers:
 - simple immobilisation of charged macromolecules
 - regeneration by washing
- Streptavidin:
 - immobilization by affinity binding
 - biotinylated substances can be immobilised (not reversible)
- His-tag:
 - Ni²⁺-surface
 - Proteins with His-tag can be immobilised (reversible)
- Membrane proteins:
 - Surface for membrane proteins
 - Immobilised using hydrogel



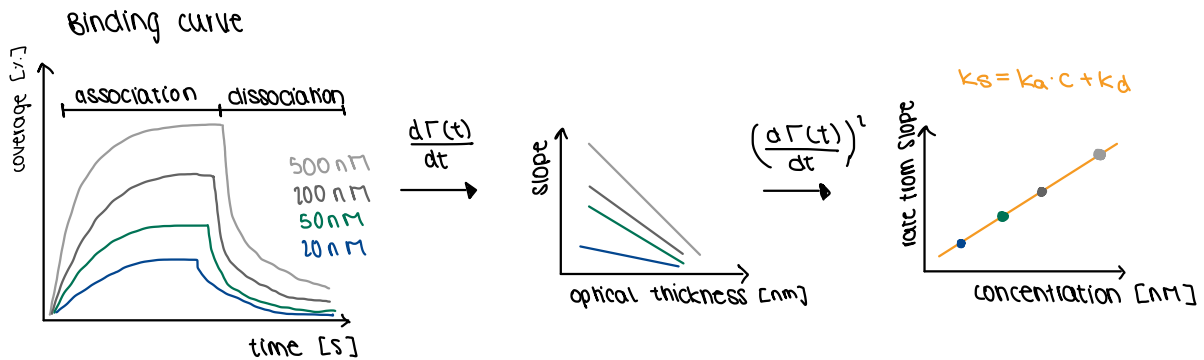
- ## Possibilities of photolinking:
- direct photoimmobilisation of biomolecules (DNA, photobiotin)
 - photofunctionalization of polymer surfaces
 - indirect photo-immobilisation using hydrogels (Photodextran)

- ## Recognition elements:
- (Functionalised) polymers
 - **Molecular Imprinted polymers (MIP)**
 - template-monomer complex is polymerised, then template is extracted
 - e.g. immobilisation of nanoparticles by electrostatic adsorption
 - ↳ advantages: many recognition sites & fast diffusion
 - Biolayers as recognition structures
 - (antigen/antibody, DNA/DNA, peptide/protein, protein/protein, small ligands (protein))
 - **Scaffolds**
 - ligand bound to stable linker scaffold
 - target protein can bind to ligand
 - e.g. Polypeptide scaffold for CRP
 - Aptamers
 - **Membrane like structures & membranes**
 - e.g. detection of salmonella
 - ↳ salmonella antigen binds to membrane-like structure & can be detected by additional salmonella antibody
 - **Nanotubes**
 - e.g. Carbon nanotubes COOH-functionalised on Amino-PEG
 - ↳ Nanoparticles can interact with nanotubes



Biomolecular interactions (BIA)

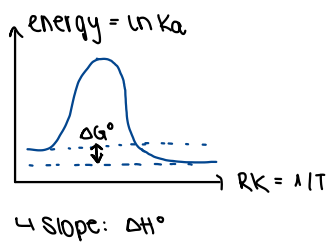
- Immuno-reactions: - analyte detection with antibodies with small molecules/proteins as analytes
- Coverage at surface depending on concentration



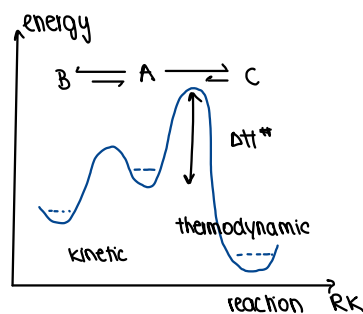
↳ Rate constant $k_s = k_a \cdot C + k_d$ is curvature of binding curve

- Overcoming problems:
 - non-specific binding → sophisticated surface chemistry
 - limit of detection → high loading
 - high throughput → imaging techniques

Thermodynamics of binding



Competitive reactions



Starting from A:

- first more B is formed
- at the end B "transformed" to C (except B is part of further fast reaction)

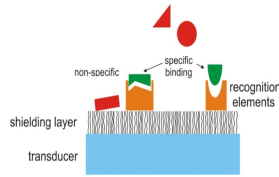
13. Optical biosensors and POCT

- Prof. G. Gausgitz

31.01.2021

Optical methods in biosensing

Biosensor principle



shielding layer:

- i) immobilisation of recognition elements
- ii) against non-specific bindings

Optical detection: Amplitude, Frequency, Phase, Polarization & Time dependence
↳ Absorption / Transmission, Emission, Scattering, Reflection & Refraction

Optical detection of biomolecular interaction in homogeneous phase

- Fluorescence quenching assay:
 - analyte derivative labelled with fluorophore
 - quenching of fluorescence by binding of ligand to receptor
- Fluorescence polarisation assay:
 - analyte derivative labelled with fluorophore
 - reduction of rotational diffusion
 - Increase in fluorescence anisotropy
- Fluorescence resonance energy transfer (FRET):
 - both analyte & receptor labelled
 - energy transfer in receptor-ligand complex
 - donor fluorescence must overlap acceptor absorption
 - ↳ Quenching of fluorescence depends strongly on the intermolecular distance
- Total internal reflection fluorescence (TIRF):
 - total reflection in transducer leads to irradiation of analytes close to surface → evanescent field
 - different analytes can be excited parallelly

Biosensor: Interaction between analyte & recognition element measured by:

- labelling of at least one partner
- without label

Optical monitoring:

- direct optical detection → reflectometry, evanescent field techniques
- detection of labelled systems → fluorescence methods, TIRF, FRET

Direct optical detection:

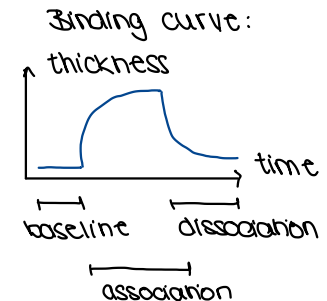
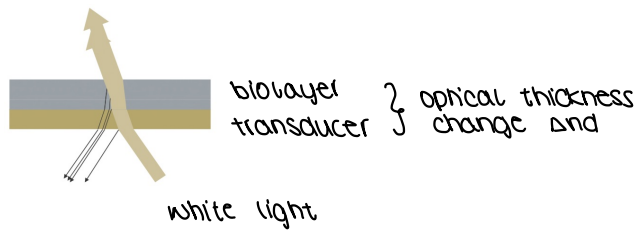
- only assays in heterogeneous phase
- requirements for interaction surface:
 - avoid non-specific binding
 - regeneration
 - stability
 - optimal specific binding
 - high loading
- no problems with label:
 - less costs
 - less problems with bioactivity
 - fast characterisation
 - no problems with photobleaching

↳ Examples of label-free principles: Quartz microbalances, cantilevers, electrochemical methods, optical principles & combinations

- ↳ Classification:
- i) Refractometry → measurement of refractive index e.g. SPR
 - ii) Reflectometry → measurement of optical thickness e.g. RIFS
- ↳ Optical thickness = $n \cdot d$ → both dependent on temperature
→ n : volume dependent polarisation

• Reflectometric interference spectroscopy (RIFS):

- neglectable temperature effect on optical thickness



Biosensor applications

- Point-of-Care Diagnostics (POCT):
 - laboratory analyses near patient (bedside)
 - performed outside laboratory
 - no sample pretreatments
 - no pipetting steps
 - use of ready-to-use reagents
 - purpose-made measuring devices for single measurements
 - operation by non-technicians
 - rapid available results
 - ↳ allow immediate therapeutic responses
- ↳ fields of application: ICU/operation room, ER, ambulance, emergency ambulance vehicle, Doctor's office, patients home
- ↳ types of POCT devices:
 - unit-use devices → sensors integrated in test strips & are to be inserted in reader instrument
 - e.g. Glucose test via amperometric detection
 - e.g. iSTAT handheld system to measure various biomarkers
 - e.g. lateral flow assays via capillary flow
 - Benchtop instruments → use various analytical principles
 - continuous monitoring devices → e.g. continuous glucose monitoring e.g. microdialysis
- ↳ Clinical perspectives for POCT:
 - intensive care needs: blood gases, coagulation, troponin, hemoglobin, electrolytes
 - hospital applications for rapid assessment of life-threatening crisis → "golden hour"
 - home care POCT supporting compliance with therapy
 - Doctor's office
 - ↳ increasing relevance in the future with trend of miniaturisation & multiplexing
 - "E-health" = citizen & data protection
 - + medical care & products/service reliability

• New technologies & analytical devices:

- Trends to use novel:
 - Biological recognition elements
 - Optical signal generation techniques
 - Magnetoresistance
 - Arrays for infectious agents, sepsis, tumors, autoantibodies, ...
- C-chip & Reader: generic microcapillary driven chip featuring covalent surface chemistry to replace conventional lateral flow membranes

Total costs per test of POCT may vary between 2-10 times the costs in the core lab

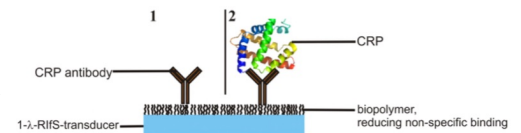
↳ but time of POCT is shorter compared to core lab

↳ Question "What is clinically useful?" must be seen for POCT applications

• Applications:

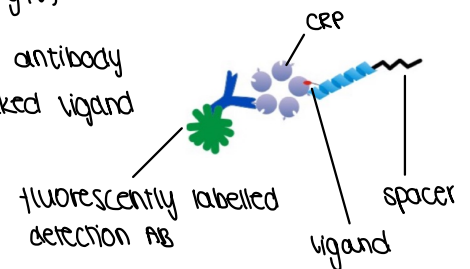
i) CRP as general parameter for inflammation

- CRP increases in case of infection
- direct optical detection is used, because big molecule
- challenge: antibodies as recognition elements on the sensor surface
 - ↳ immobilization is problematic
- LOD is small enough to sensitively detect CRP with RfIS (0.32 mg/l)



- ↳ new approach:
 - Sandwich TIRF with fluorescently labelled detection antibody
 - capture element: peptide scaffold with independently linked ligand & spacer

↳ working range can be shifted to lower concentrations (0.04-2.3 mg/l)



ii) Neopterin assay

- Neopterin is marker for HIV infection, CVD, organ transplantation, autoimmune disease, tumors, T-lymphocytes activity
- small molecule, therefore binding inhibition test is used
- LOD & LOQ sufficient

iii) Antiphospholipid syndrome

- = auto-immune disease leading to dysfunction in blood coagulation
- β_2 -glycoprotein I (β_2 GP-I) is main auto-antigen
- Transducer coated with model membrane
 - ↳ β_2 GP-I is attached to membrane leading to reactive conformation
 - ↳ Biomimetic approach mimicking the disease on the transducer & label-free detection (RfIS)
- ↳ Essential for emergency diagnostics during pregnancy

iv) Cystatin-C

- marker for renal function, better than creatinine as creatinine is strongly influenced by external factors
 - ↳ increase in serum level indicates kidney dysfunction
- Binding inhibition assay used → small protein

V) Detection of acute pancreatic failure

- Biomarker enzyme lipase in blood
- Binding inhibition assay is used, regeneration of surface possible

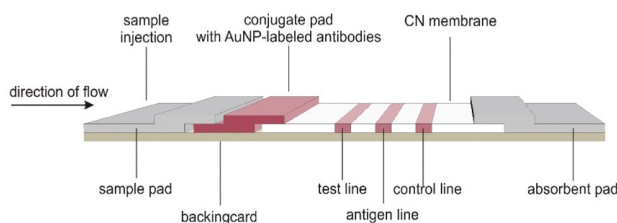
VI) Therapeutic drug monitoring (TDM)

- Inter- & intra-patient variability related to pharmacokinetics & blood level
- Control of blood level influenced by some pharmaceuticals
- faster/better response to therapy
- Reduction of spillover effects
- Safety → avoidance of toxic concentrations
- Cost efficiency

VII) Application of amitriptylin

- AMT was one of the first anti-depressiva with nortriptylin (NRT) as main metabolite
- only very small "therapeutic window": 0.68 - 11.99 mg/l
- Binding inhibition test with RFIS used to determine concentration

VIII) Lateral flow for medical diagnostics



- antigen line to hinder that non-blocked labelled antibodies are lost
- ↳ ratio between test & antigen line gives better information

- ↳ several modifications available e.g. gold nanoparticles to label antibodies
- e.g. multi-channel chip to analyze multiple biomarkers

- ↳ problem: calibration of lines as they are not made for quantification

IX) Parallelized Peptide-Antibody Interactions

e.g. Epitope mapping for anti-Transglutaminase Antibody in celiac disease

- different interaction behaviour between 5 sequence consecutive peptides & a monoclonal AB
- ↳ choose "best peptide"
- binding signal dependent on antibody concentration
- transition between mass controlled & kinetically controlled conditions

X) Nanoderm for therapeutic drug monitoring

14. Metabolomics

- Dr. C. Trautwein

01.01.2021

I. Metabolomics

= Study of main pathways of human metabolism & metabolites (< 1000 Da → "small molecules")

↳ Downstream all other big "omics" spheres

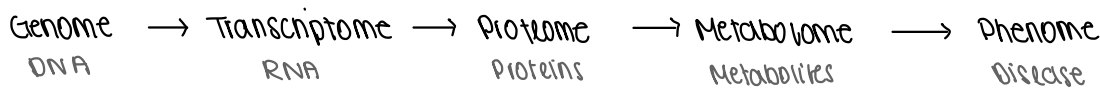
e.g. DNA → genomics, RNA → transcriptomics, proteins → proteomics, metabolites → metabolomics

• Relevant, because small molecules involved in food flavors & aromas, common clinical tests, drugs, most common diseases & natural compounds

↳ Lifetime risk to have diseases related to "small molecules", e.g. cold, influenza, hypertension, cancer, obesity, diabetes, stroke, ...

• Metabolites can derive from human cells, nutrition, microbiota & exposures ("exposome")

• Definition Metabolomics: - represent functional endpoint of physiological & pathophysiological processes
- depict both genetic predisposition & environmental influences

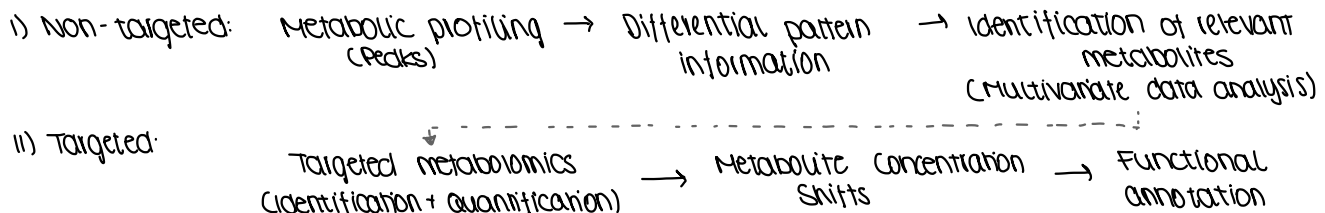


• Goals in modern clinical diagnostics: - early diagnosis - subtyping/staging
- prophylaxis instead of therapy - therapeutic drug monitoring
- phenotypic pharmacogenomics - individualized

↳ Paradigm shift in clinical diagnosis: High-resolution biochemical characterization of body fluids by targeting metabolomics will facilitate transition
- from damage-oriented to function-oriented diagnostics
- from risk factor assessment to actual pathophysiological phenotyping
- from generic diagnostics of complex diseases to a detailed sub-typing & staging

↳ Find & identify medically useful biomarkers

• Strategic approaches:



II. Techniques

To date, no single analytical technique is able to detect all metabolites within a sample.

This complicates analysis.

1) Mass Spectrometry: - Gas chromatography MS (GC-MS) } 100-
- Liquid chromatography MS (LC-MS) } 400 k€

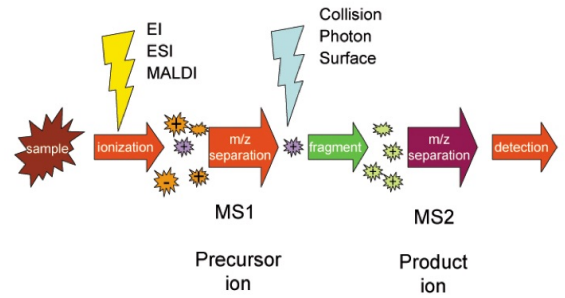
↳ consists of Ion Source, Mass Analyser & Detector

e.g. Time-Of-flight, Triple quadrupole, Linear ion trap

4 Tandem mass spectrometry:

- Precursor ("parent") ion
- Product ("daughter") ion
- Neutral loss scan
- Selected reaction monitoring

e.g. MALDI imaging of breast cancer margin



(I) Nuclear Magnetic Resonance Spectroscopy (NMR) → 1.5-2 Mio. €

→ use of spin echoes, possible for nuclei with spin $\neq 0$, e.g. ^1H , ^{13}C , ^{19}F

→ Free Induction decay (FID) $\xrightarrow{\text{FT}}$ Spectrum

- examples of 1D NMR:
- high fat & normal food pellet extract
 - glycerol peak visible
 - joint arthritis polar extract
 - plasma polar extract

→ examples of 2D NMR: - COSY Spectrum of *Bacillus* sp. 4115 overlaid with spectrum from M1 medium

(II) Magnetic Resonance Mass Spectrometry (MRMS)

→ Overlap of detected metabolites by principal metabolomic techniques
e.g. Human urine & serum metabolome

III. Samples

- **Biofluids:** e.g. urine, plasma/serum, saliva, synovial fluid → can be directly measured
- **Tissues, organs & cell cultures:** e.g. nervous tissue, gall bladder → **extraction procedure needed**
- **Metabolites have very fast turnover:** e.g. mammalian cells: metabolites $\sim 1\text{min}$
mRNA $\sim 10\text{h}$
proteins $\sim 1\text{day}$

→ proper workflow is crucial:

sample collection → sample preparation → sample analysis → data acquisition → data analysis & interpretation

Quenching

- **Reliable quality control markers in metabolomics experiments is essential**
 - Adenylate Energy charge → Quenching efficacy
 - Redox Status (GSH & GSSG)
 - Contaminations

IV. Examples

- **Heatmaps:**
 - single gene knockout in cell culture → very clear result
 - antibiotics treatment in mice → almost clear result
 - Alzheimer's disease in humans → not clear at all
- **cell cultures & preclinical samples:**
 - cell culture PINK1 knockout
 - mouse microbiome
 - mouse inflammation

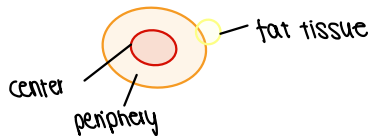
complexity & external influences

- Human glioma tissue cells: detection of predictive markers for molecular tumor board
 - ↳ translation: MRI → MRS → NMR
 - ↳ PCA to find biomarker hypotaurine

• Human breast cancer tissue samples:

- I) Multicellular organisms → glucose uptake depending on growth signal
- II) Differentiated tissue → oxidative phosphorylation or anaerobic glycolysis depending on O_2 level
- III) Proliferative tissue → Anaerobic glycolysis not dependent on O_2 level
 - ↳ Warburg effect in oncology (tumor metabolism)

→ Breast cancer tumor heterogeneity



Luminal A: ER + PR + HER2 -

Luminal B: ER + PR - HER2 +

HER-2: ER - PR -

Chemotherapy with trastuzumab

ER = estrogen receptor, PR = progesterone receptor

HER2 = human epithelial growth factor

→ workflow of deep-frozen tissue:

1. Deep-frozen tissue is cryogenically pulverized & lipophilized
2. Samples extracted with 2-phase liquid: liquid extraction
3. Phases are separated, evaporated to dryness & reconstituted with deuterated buffer
 - ↳ Performance of metabolomics & RNA analysis reveal different factors in different zones of breast cancer

→ ascorbate has been proven to destroy tumor cells in center region by rewiring the metabolism via redox imbalance & energy crisis

- Intraoperative analysis: e.g. analysis of cancer metastases in lymph nodes

V. Personalized Medicine

- Metabolomics represents functional end-point of physiological & pathophysiological processes depicting both genetic predisposition & environmental influences like nutrition, exercise & medication
 - ↳ Approach: Quantitative metabolomics to define health & disease
- Many diseases are mixture of predispositions
 - ↳ Precision medicine needs to be quantitative & will benefit from multi-scale sampling
- Human samples require biobank QC & strict SOPs → NMR as promising analysis device
- Identification of disease-specific metabolite markers
 - ↳ Paradigm shift in clinical diagnosis: conventional diagnosis → multiparametric diagnosis
 - ↳ What concentrations are healthy? Which depict physiological processes?
- Summary:
 - Metabolomics could be future add-on to existing lab values
 - Quantitative reference databases need permanent input
 - SOPs need to be standardized, independent on NMR/MS
 - Use of disease specific panels instead of single metabolites

- add other omics levels (exome, epigenome, transcriptome)
- Screening of high risk patients
- Perform monitoring for treatment success & relapses
- **Examples:**
 - Metachromatic Leucodystrophy: Detection of Neopterin as biomarker metabolite
 - SARS COV-2 : infection paralyzes cytotoxic & metabolic functions of immune cells

VI. Summary

- Metabolomics is established omics research field
 - true, when applying simple lab experiments with model organism (cells, mice, ...)
 - not true, when dealing with real human samples
 - ↳ strict SOP & biobanking AC needed with measurements being quantitative & reproducible
- Different to NGS : techniques for metabolomics still have not reached a final technical level
 - in MS many techniques compete (qTOF, Orbitrap, TQF, ...)
 - for NMR 600 MHz has been set as gold-standard
 - ↳ combination of MS & NMR based techniques
- Metabolomics have great potential for personalized precision medicine
 - use of predictive metabolite panels for early stage diagnosis
 - Multi-omics data integration helps to better understand phenotypes
 - Longitudinal sampling will allow for lifetime trajectories & treatment monitoring
 - Provide measures for health & disease