# Protein Engineering

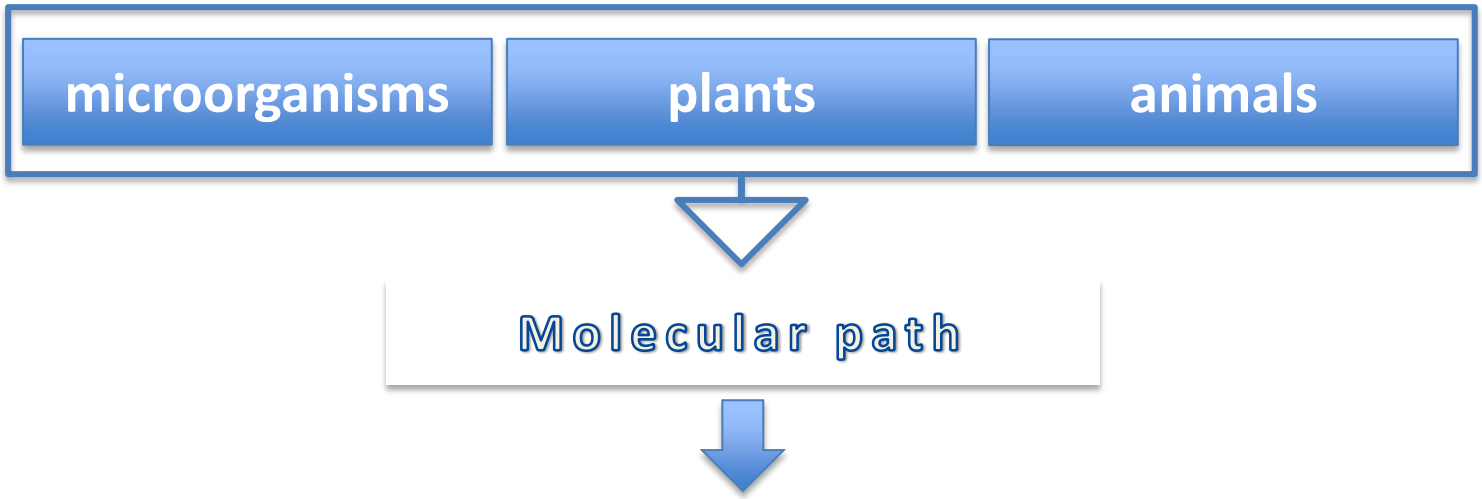
# Enzyme Engineering

# Molecular Biotechnology - Biocatalysis

- **Access to a broad diversity of biocatalysts**
  - -- natural diversity → „GENOMICS“
  - -- artificial diversity → „SYNBIO“
- **Economic production of enzymes**
  - -- recombinant enzymes
- **Efficient biocatalysts for any application**

  (fast and efficient methods for the development of enzymes)
  - -- enzyme engineering → “DIRECTED EVOLUTION“
    → “RATIONAL DESIGN“
- **Novel biocatalysts**
  - -- Nanobiotechnology
  - -- novel catalytic structures

# Recruitment of enzymes from natural biodiversity



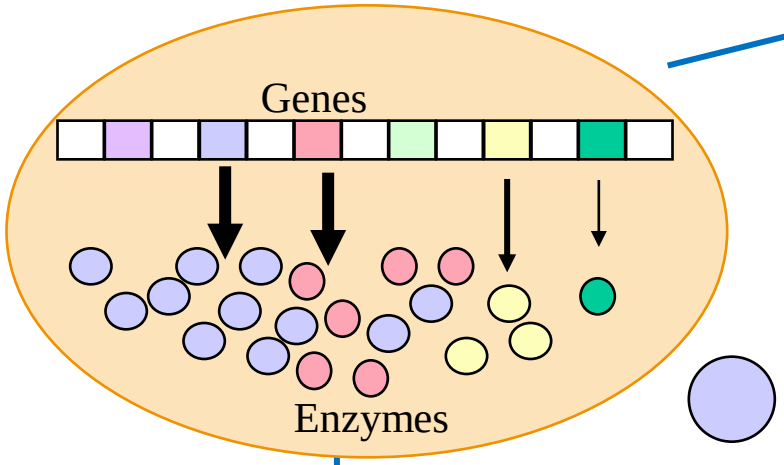

Isolation and cloning of gene(tic) material

Identification of clones with specific enzyme activity

**Gene expression – gene technological production of enzymes**

# Recruitment of enzymes from natural biodiversity

Isolated Organism



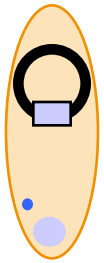

Screening

Natural diversity

Organism producing an interesting enzyme was identified

Isolation, purification and characterization of the enzyme

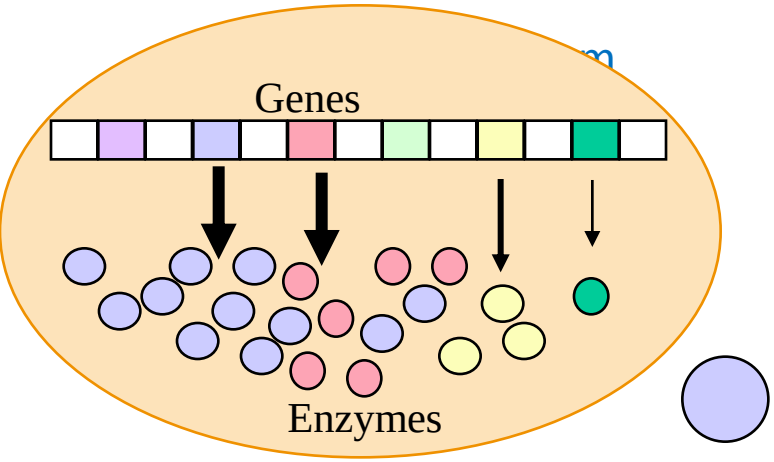
Cloning and expression of the specific gene



Recombinant microorganism

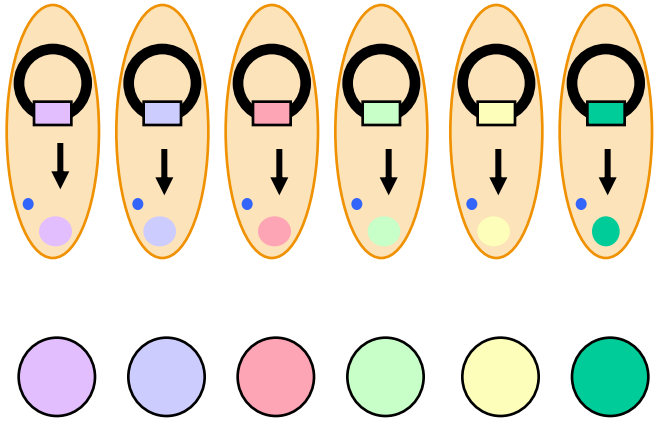
Recombinant protein

# Recruitment of enzymes from natural biodiversity




Genomic Library

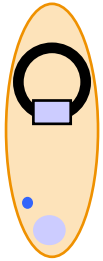
Gene Libraries



Screening

Natural diversity

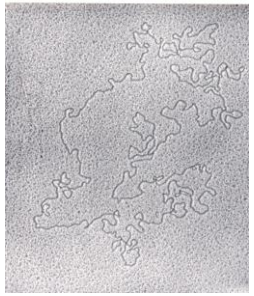
Screening



Recombinant microorganism

Recombinant protein

# Recruitment of enzymes from natural biodiversity



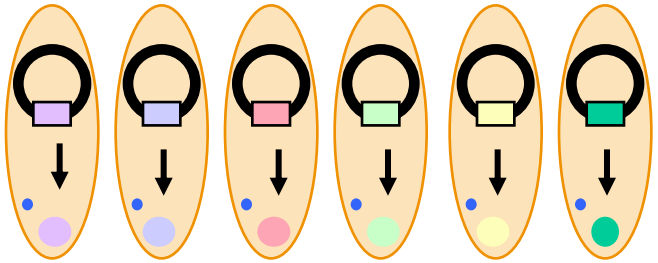
Cloning of DNA fragments

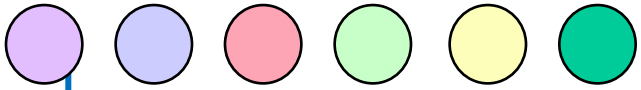
Gene libraries

Total DNA isolation

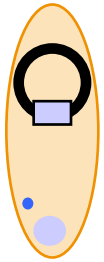
Non-cultivatable diversity

## Metagenome Library





Screening



Recombinant microorganism

Recombinant protein

# Recruitment of enzymes from natural biodiversity

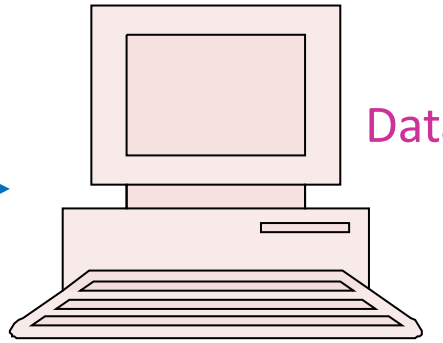
(Meta) Genome sequencing

David Parker/Science Photo Library/Photo Researchers, Inc.

**Much of the speed** with which recent advances in genetics research have been made results from the use of high-throughput DNA sequencers coupled with computerized sequence acquisition, like these devices at the Sanger Centre near Cambridge, England. This technology has made it possible to determine the complete DNA sequence of the human genome.

Natural diversity

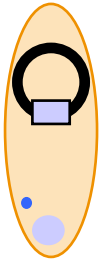
Sequence Database



Data Mining

Protein Sequence

Gene Design
Gene Synthesis



Recombinant microorganism

Recombinant protein

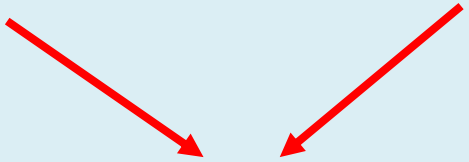
# Designed Evolution

## Concept of „Process Designed Enzymes“

**Establish set of key enzymes – e.g. esterases**
**key structures / functionalities**
**genes - expression**

**Develop efficient routes to enzyme engineering**
**tuning enzymes towards specific process needs**

**Directed Evolution – Rational Design**



**Designed Evolution**

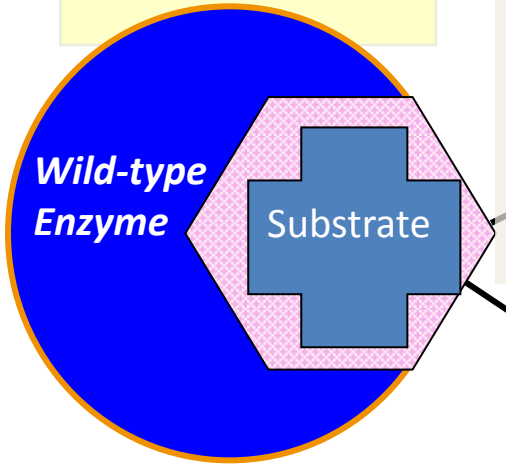
# Enzyme-Engineering → basic routes

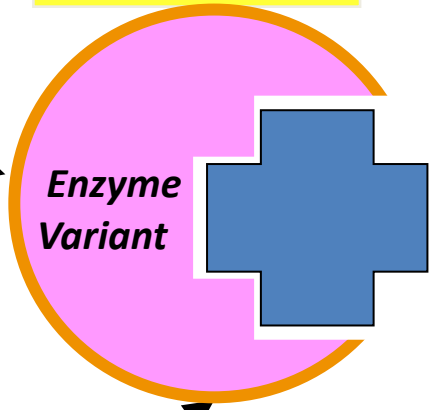
low activity
insufficient stability
unfavoured specificity
bad selectivity

**Molecular Evolution**
random mutagenesis
gene shuffling

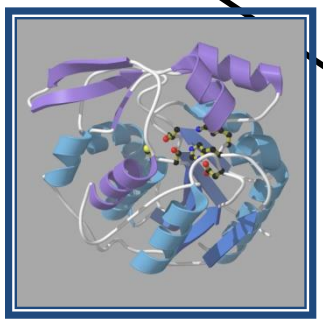
Variant library
Screening - Selection

high activity
excellent stability
proper specificity
good selectivity




**Functional and molecular Characterization**



defined specific variant
targeted construction

**Rational Engineering**
Structure-Function Relationships
Modelling, Design

# Prerequisites for Protein Engineering by Rational Design

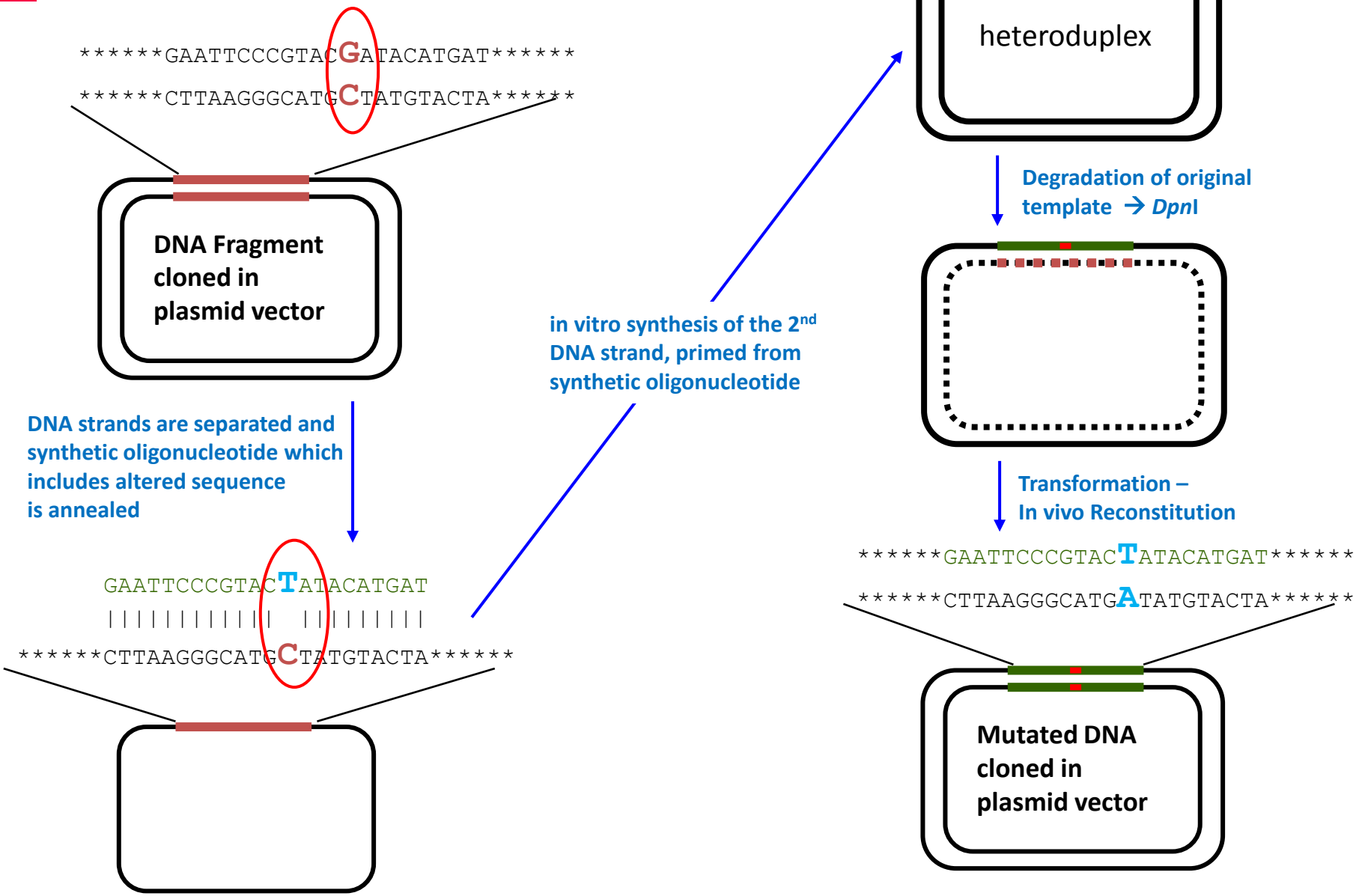
**Availability of structure information**

- **X-ray crystallography**
  - „frozen structure information“
  - need for crystallization
- **NMR structure analysis**
  - restricted to small proteins
  - information on protein dynamics possible
- **Modelling of structures based on aa sequences and homologies**
- **Modelling of substrate-Protein interactions - docking**

**Availability of information on structure-function relations**

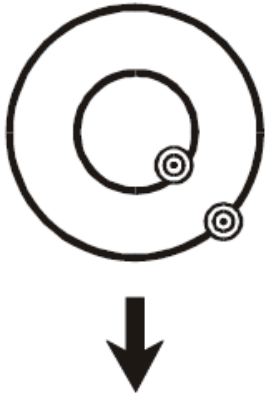
- **Information on molecular mechanisms of biological functions**
  - e.g. reaction mechanism, protein-small molecule interaction (e.g.
- **Information on aa residues involved in biological function**
  - e.g. active site residues…
  - e.g. sites for cofactor binding …

# Site Specific Mutagenesis



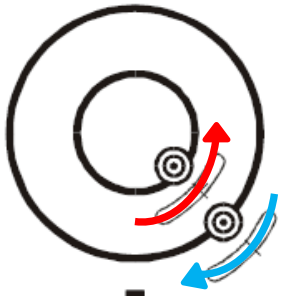
# QuickChange™ Mutagenesis System

**Step 1**
**Plasmid preparation**

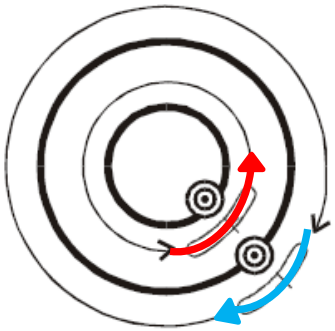


**Gene in plasmid with target site for mutation**

**Step 2**
**Temperature Cycling**

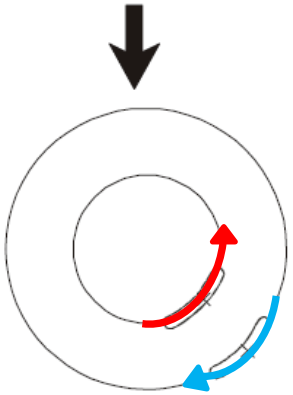


**Plasmid denaturation**
**Annealing of oligonucleotid primers containing the desired mutation**



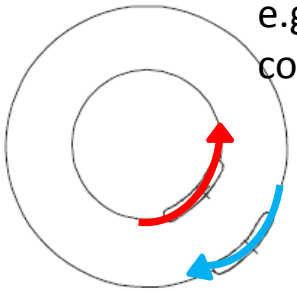
Extension and incorporation of mutagenic primers by *PfuTurbo* DNA Polymerase resulting in nicked circular strands

**Step 3**
**Digestion**



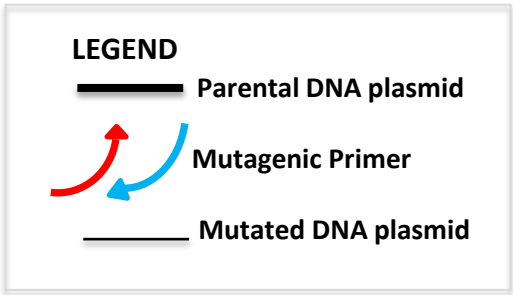
DpnI digestion of the non-mutated, parental DNA template

**Step 4**
**Transformation**



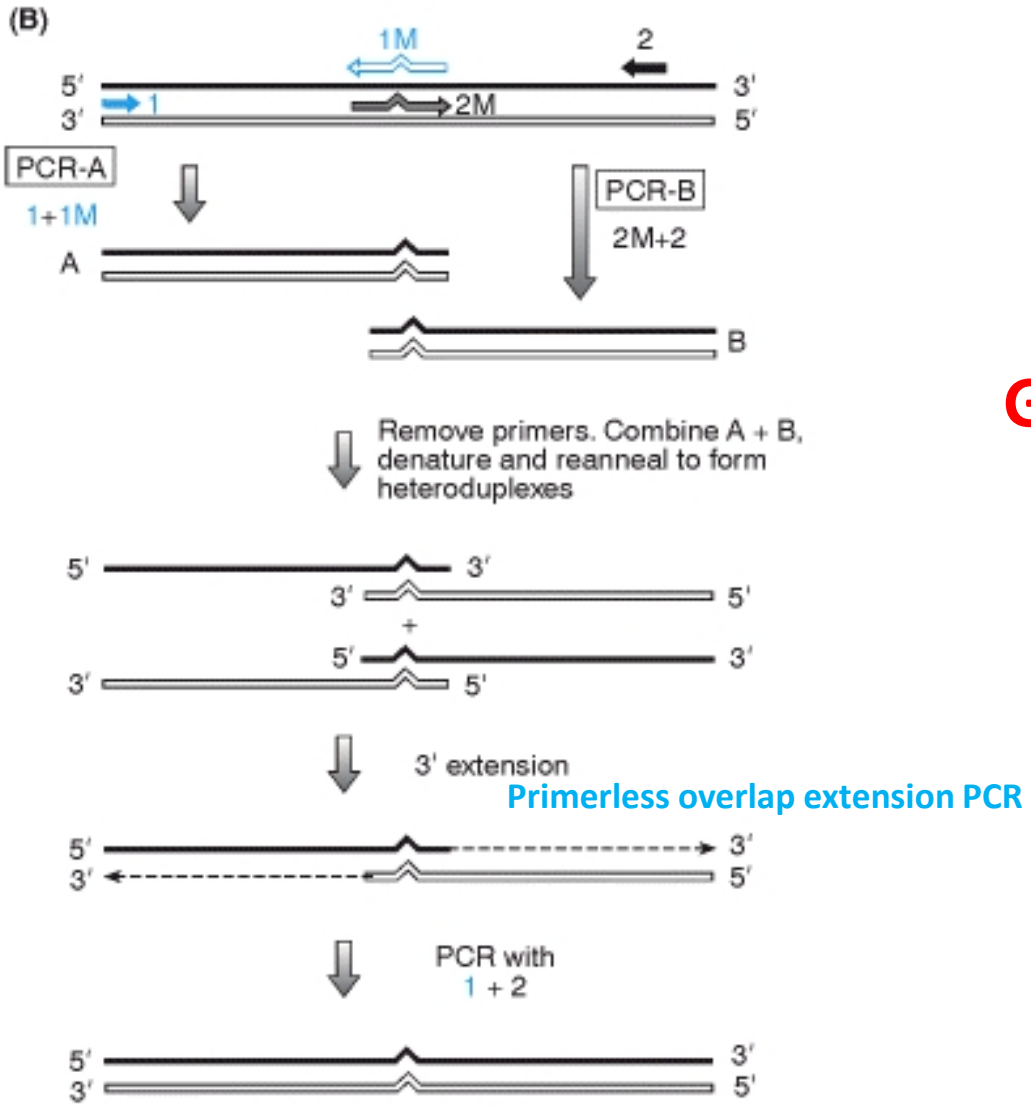
Transformation of the circular, nicked dsDNA into e.g. *E.coli* XL1-Blue competent cells

***E.coli* cells repair the nicks in the mutated plasmid**




**Taken from: Stratagene QuikChange Manual**

# PCR- mediated, site directed Mutagenesis




## Gene synthesis

Primerless overlap extension PCR

Taken from: Human Molecular Genetics. 2nd edition. Strachan T, Read AP. New York: Wiley-Liss; 1999.

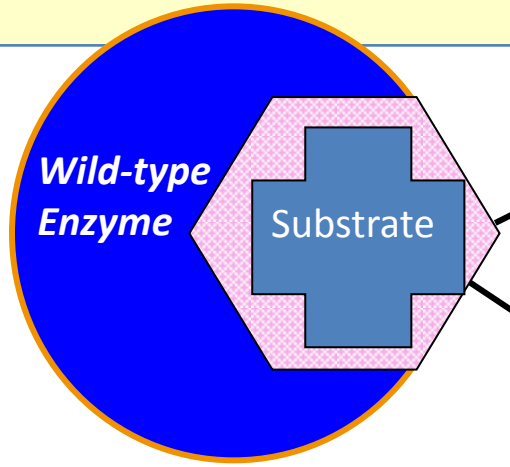
23.4.15

# Enzyme-Engineering → basic routes

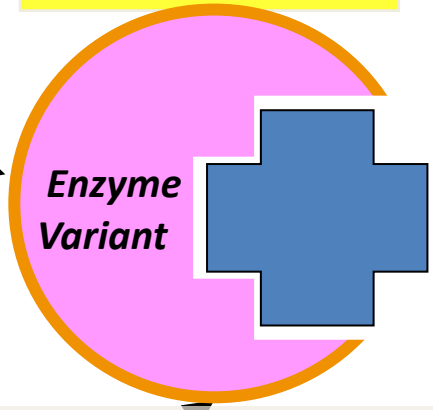
low activity
insufficient stability
unfavoured specificity
bad selectivity

**Molecular Evolution**
random mutagenesis
gene shuffling

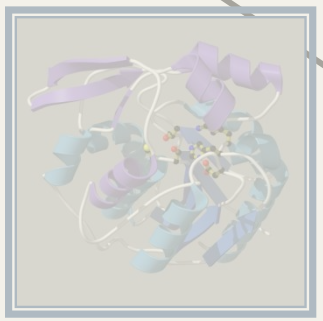
high activity
excellent stability
proper specificity
good selectivity




Variant library
Screening - Selection



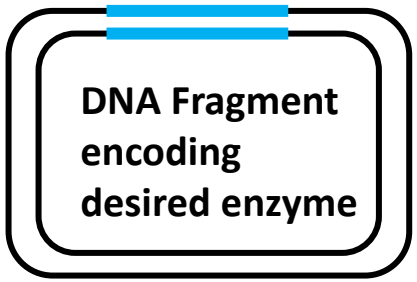

Functional and molecular Characterization



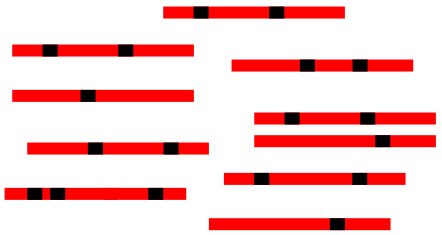
defined specific variant
targeted construction

**Rational Engineering**
Structure-Function Relationships
Modelling, Design

# Directed Evolution - outline

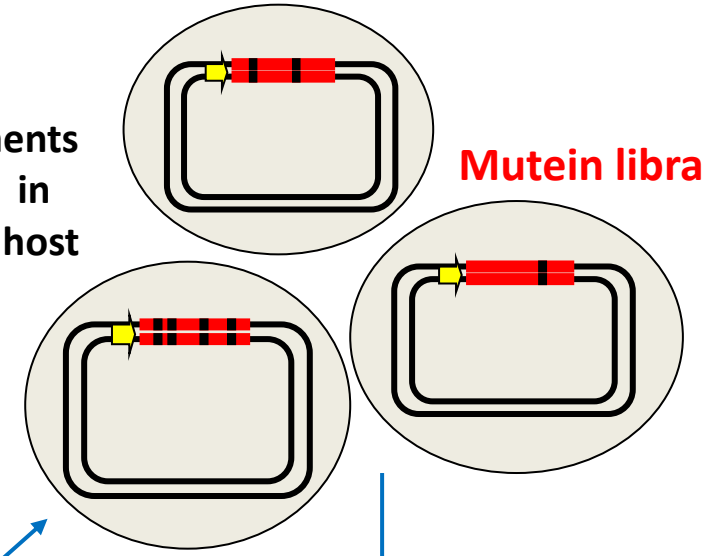



DNA fragments
derived by synthesis
PCR synthetic gene



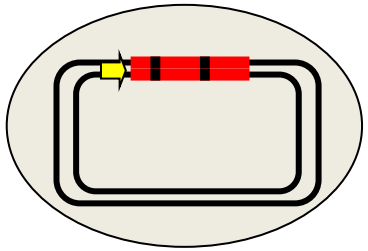
Generation of artificial diversity
Random Mutagenesis
Gene Shuffling

DNA Fragments
Introduced in
expression host



Mutein library

Screening
Selection

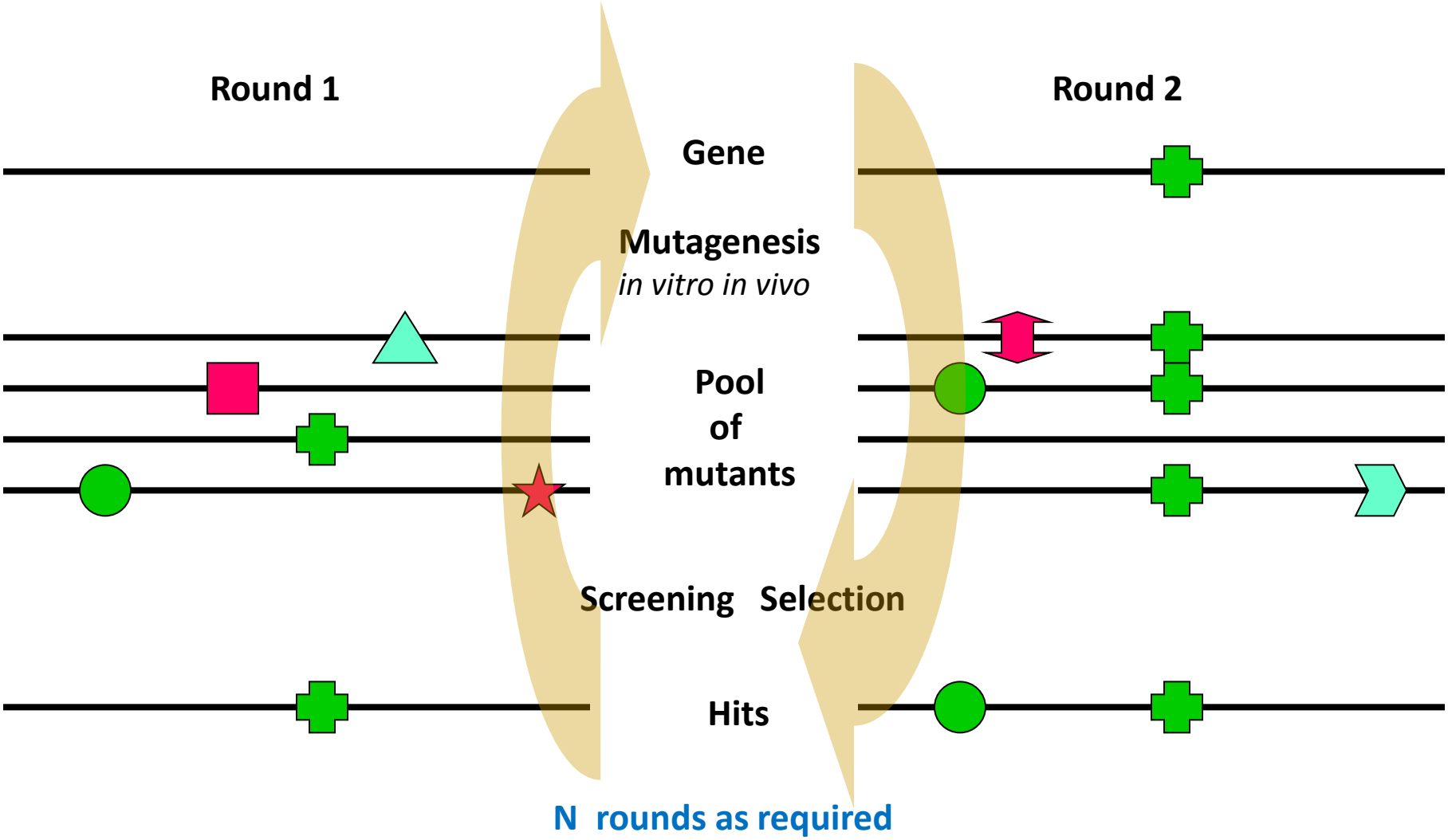


Hits

Cultivation
Functional *in vivo* expression
of enzyme variants

# Molecular Evolution of Enzymes

## Random mutagenesis



# Generation of Mutant Libraries

## Random mutagenesis of entire coding region

- error prone PCR, SeSaM
- *in vivo* mutation systems (mutator strains, transposons)
- deletion and insertion strategies (scanning mutagenesis)

## Random mutagenesis of selected parts of coding region

- cassette mutagenesis with degenerate oligonucleotides
- megaprimer PCR

## Site Saturation Mutagenesis

- All possible amino acids at specific position(s)

# Mutagenesis by Error prone PCR

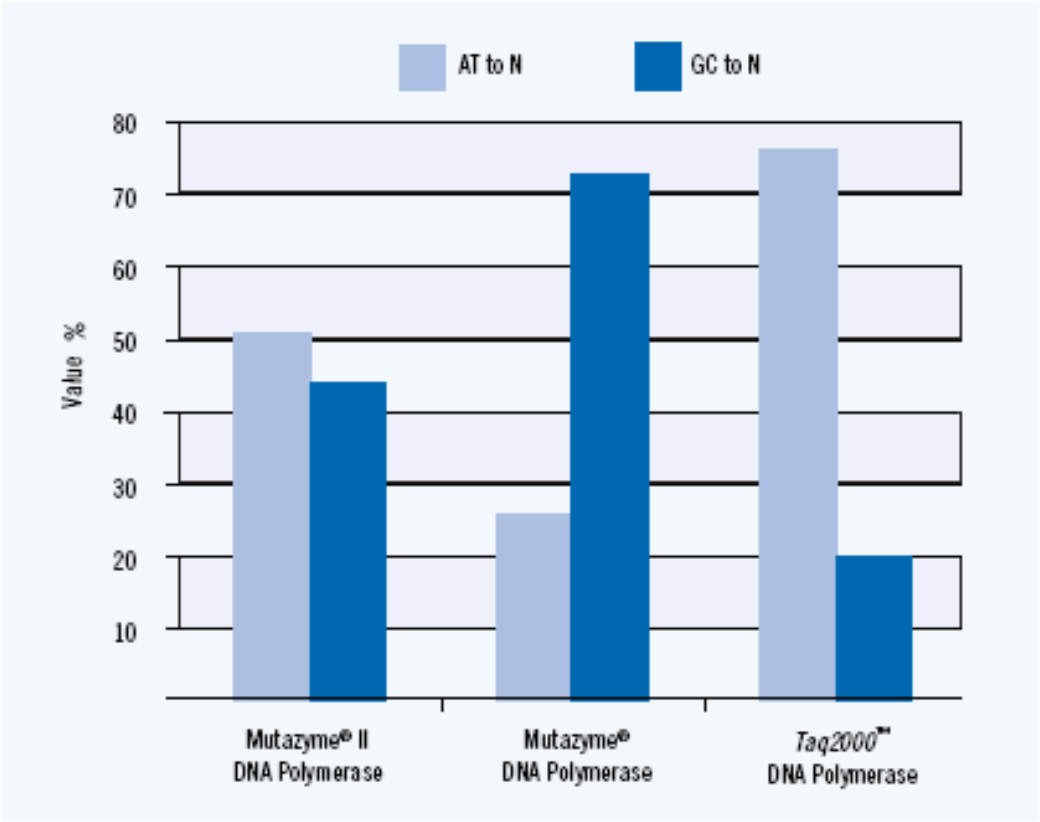
Error Prone PCR – mutagenic conditions

↑

$[Mg^{++}] \leftarrow\rightarrow [Mn^{++}]$

[dNTP]

*Taq* Polymerase → Mutazyme




**_Taq_ DNA Polymerase**

Bias for mutating A and T under error-prone conditions,

**Mutazyme® DNA polymerase**:

Bias for mutating G and C

**Mutazyme II DNA polymerase**:

Blend of Mutazyme polymerase and a novel *Taq* mutant → exhibits a higher error rate

STRATAGENE

# Error Prone PCR - conditions

S A B C D F

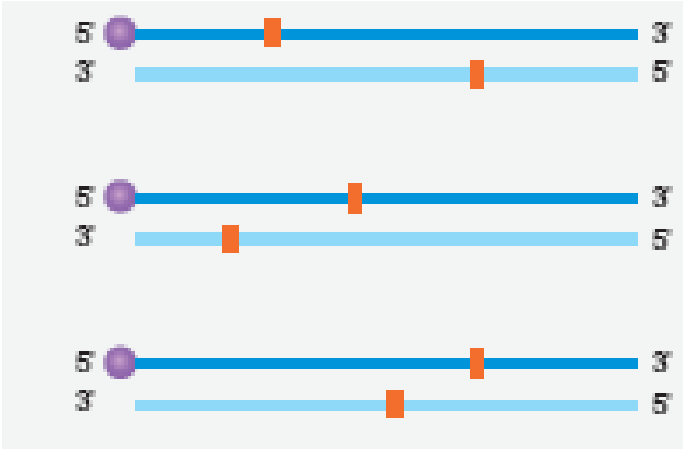
| Condition | $[MgCl_2]$ | $[MnCl_2]$ | [dNTPs] |
|---|---|---|---|
| Standard | 1,5mM | | 0,1mM each |
| A | 7mM | | 1mM each |
| B | 7mM | 0,2mM | 1mM each |
| C | 7mM | 0,5mM | 1mM each |
| D | 7mM | 1,0mM | 1mM each |
| F | 7mM | 0,5mM | 1,0mM dCTP + dTTP each<br>0.2 mM dATP + dGTP each |

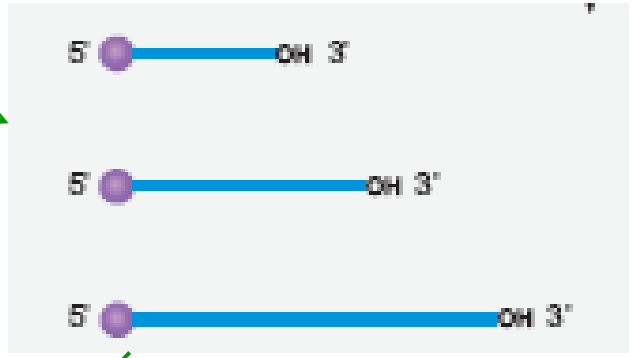
| Conditon | sequenced basepairs | A -> C<br>T -> G | A -> G<br>T -> C | G -> A<br>C -> T | G -> T<br>C -> A | A -> T<br>T -> A | G -> C<br>C -> G | Insertion<br>Deletion | Mutation rate |
|---|---|---|---|---|---|---|---|---|---|
| Standard | 2388 | | | | | | | - | 0,00% |
| A | 2388 | 1 | 1 | | | 2 | | - | 0,17% |
| B | 2388 | | 7 | 2 | | | | - | 0,34% |
| C | 2388 | | 10 | 2 | 1 | 4 | 1 | - | 0,76% |
| D | 2388 | | 10 | 1 | 1 | 4 | 3 | - | 0,85% |
| F | 2388 | | 2 | 2 | | | | - | 0,17% |

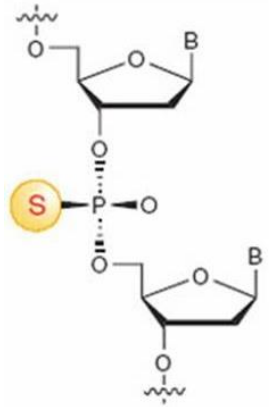
# Mutagenesis by Sequence Saturation Mutagenesis



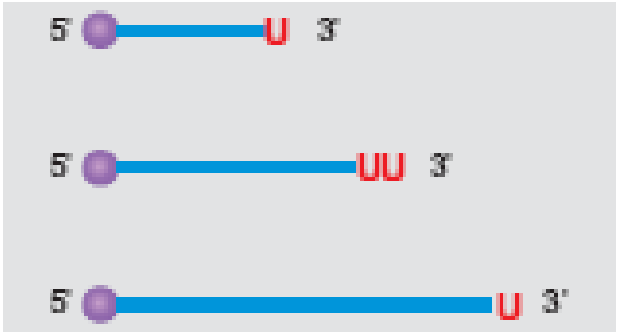

PCR amplification with biotinylated primer in presence of thiophosphate-dNTP




Cleavage at thiophosphate-dNTP positions with iodine under alkaline conditions



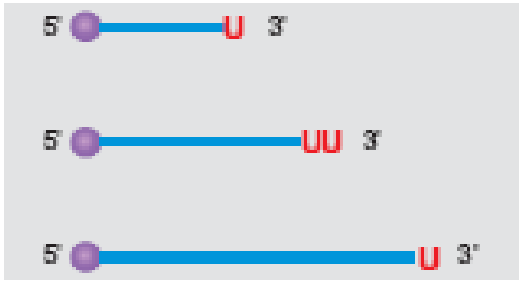

Addition of dUTP using terminal transferase, thereby at one position several nucleotides can be incorporated giving the chance to produce any triplett

SeSaM

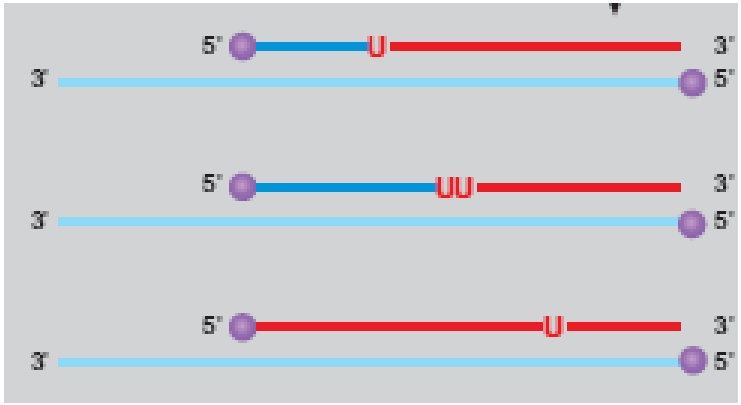
Schenk et al. (2006) Biospektrum 3, 277

# Mutagenesis by Sequence Saturation Mutagenesis

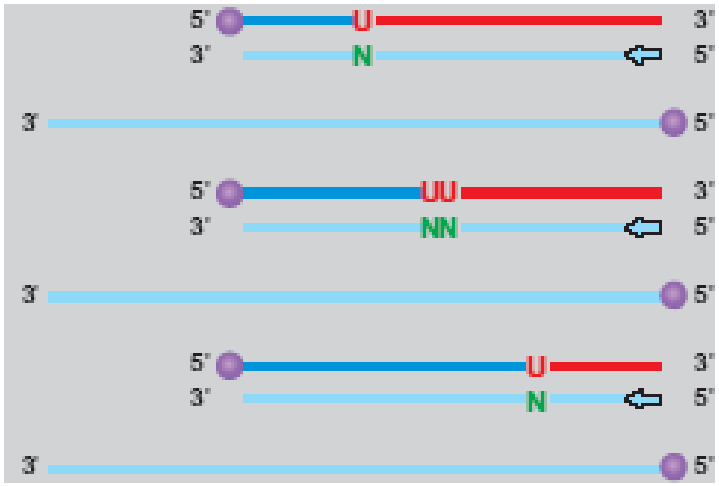




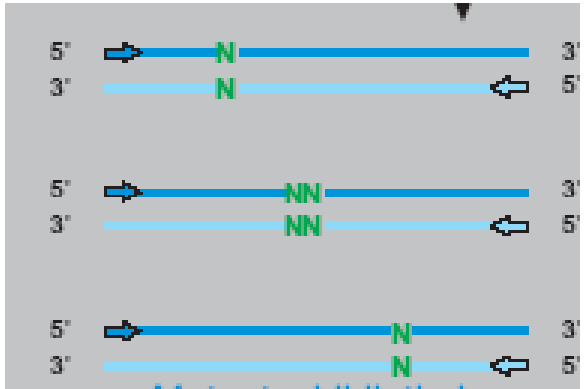

Annealing of wild type strand to biotinylated template and elongation




2nd strand synthesis with dNTPs




PCR amplification




Schenk et al. (2006) Biospektrum 3, 277

**SeSaM**

# *Experimental Library Limits*

**Length N , M aa => $M^N$ aa sequences**

| | N | $M^N$ | Mass of Library |
|---|---|---|---|
| | 3 | $10^4$ | |
| | 5 | $10^6$ | |
| | 10 | $10^{13}$ | Milligrams |
| | 20 | $10^{26}$ | Tons |
| | 50 | $10^{65}$ | Mass of Earth |
| | 100 | $10^{130}$ | |
| Typical Protein Size | 200 | $10^{260}$ | |

2.20.2000

XENCOR

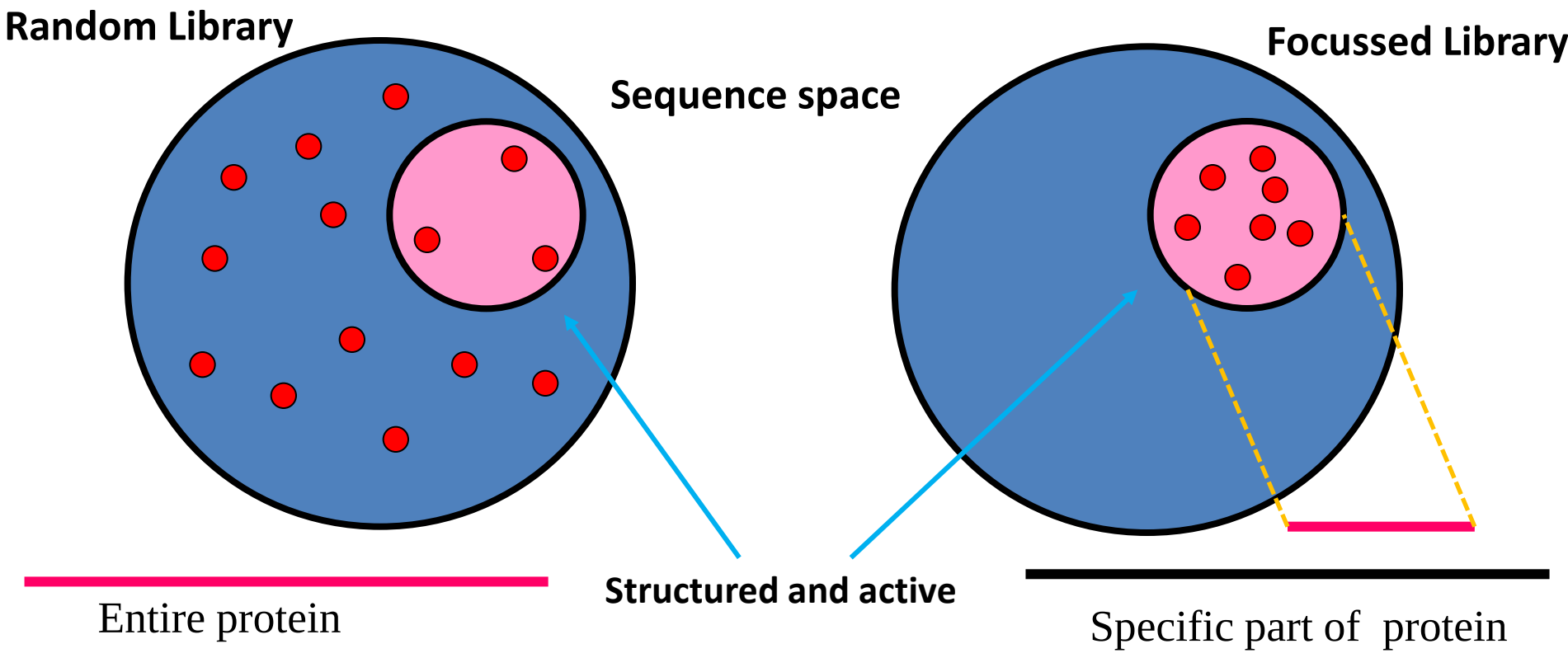
# Library management



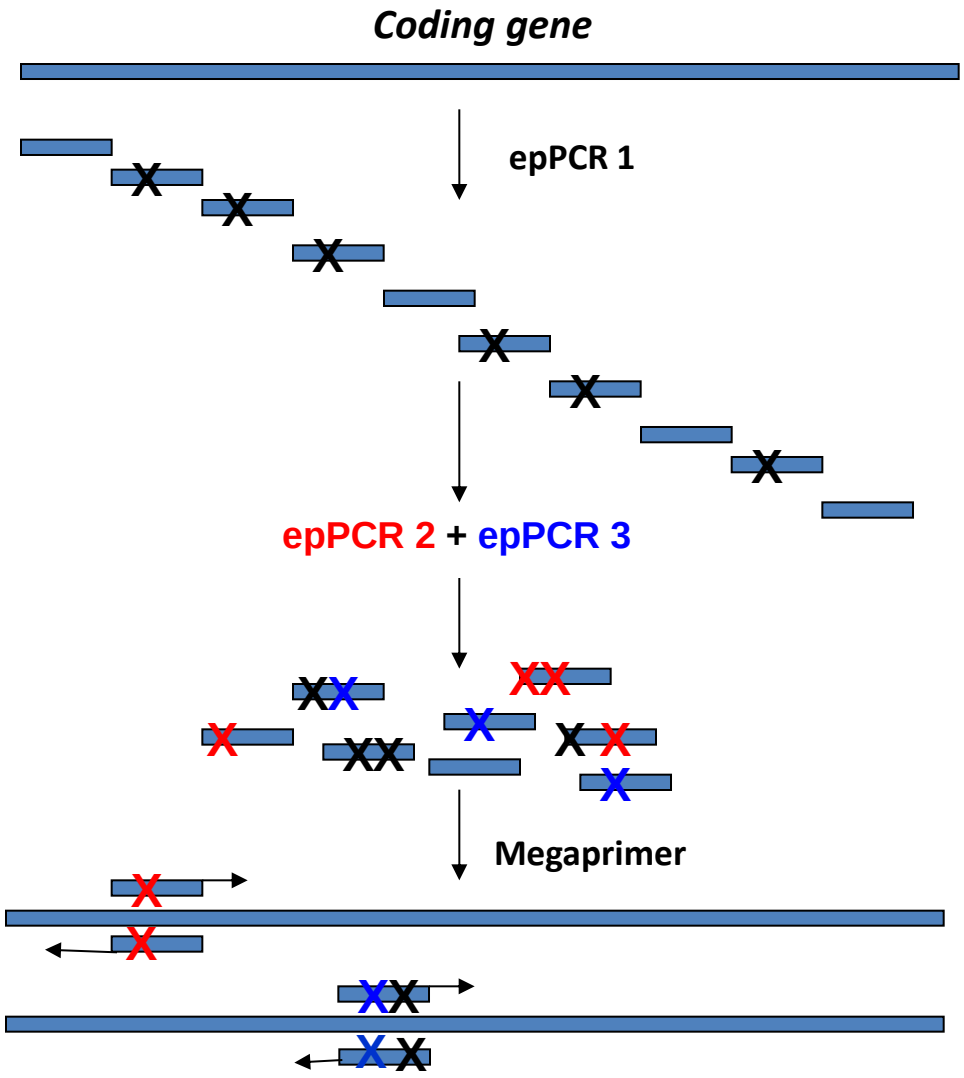
Numbers of possible protein variants (Kuchner and Arnold, 1997)

$$V = (19)^M \times \frac{N!}{(N-M)! \times M!}$$

| Number of aa changed simultaneously (M) | Sequence length (N) | |
|---|---|---|
| | 5 | 477 |
| | Number of possible variants | |
| 1 | 95 | 9063 |
| 2 | 3610 | 40982886 |
| 3 | 68590 | $1.239 \times 10^{11}$ |



# Library management
# Partial fragment mutagenesis



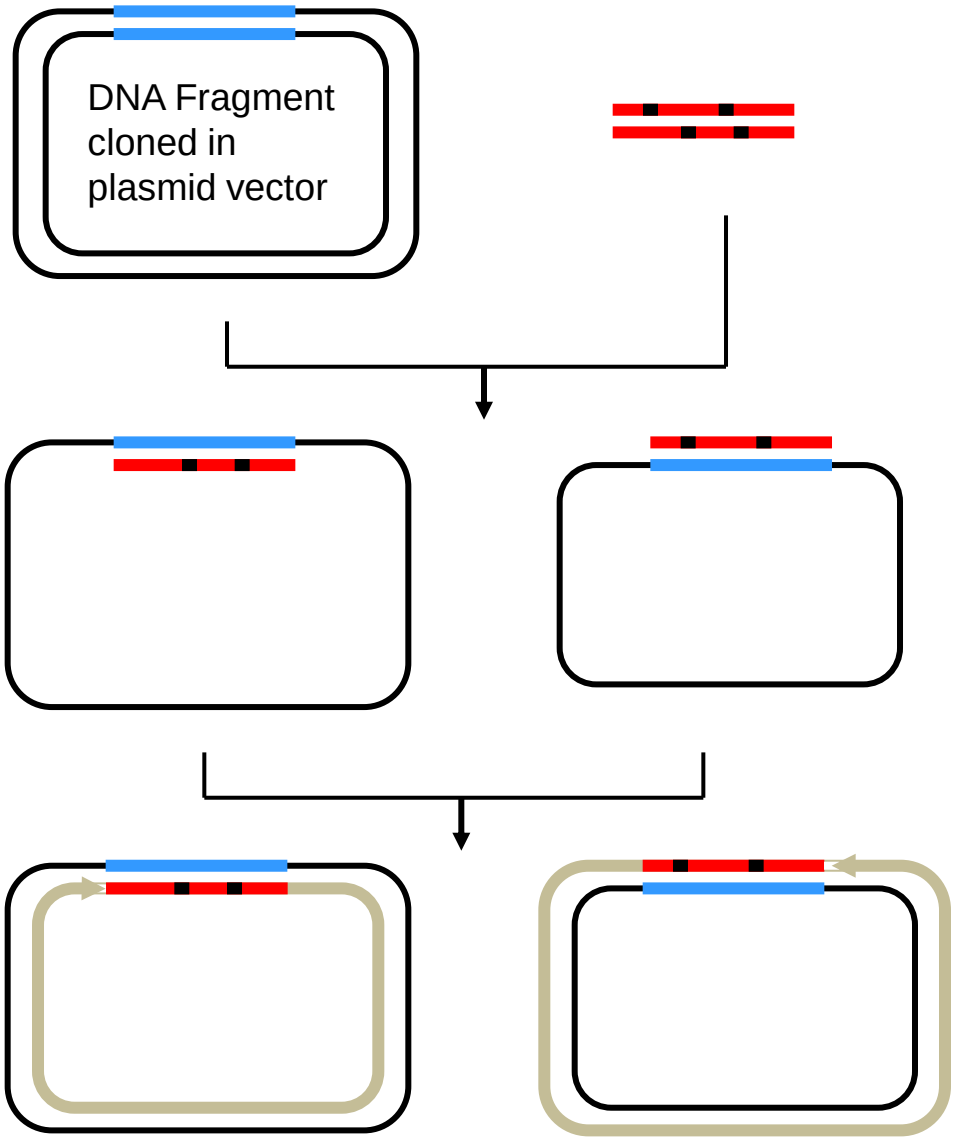

⇨Mutagenesis of subfragments

⇨DNA from epPCR 1 used as template for epPCR 2, etc…

⇨ Mutagenized fragments are introduced in expression vector by megaprimer PCR

⇨ 1- 3 mutations per ~100bp (fragment)

# Random Mutant Libraries by Megaprimer PCR



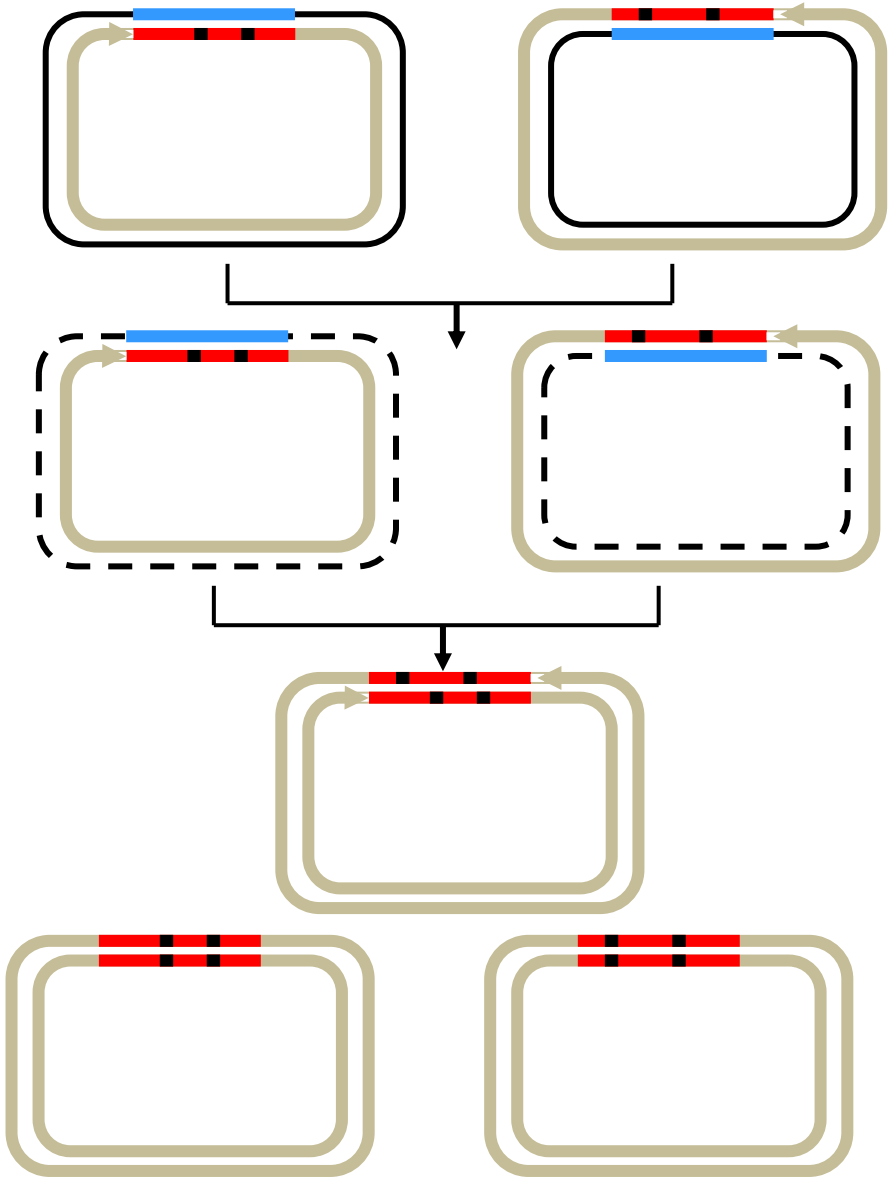

**Mutagenize DNA fragment**
* **PCR**
* **Degenerate gene synthesis**
→ **Megaprimer**

**Denaturing**
**Megaprimer annealing**

**Extension with**
**DNA polymerase**
**(PCR)**

*K. Miyazaki (2003)*
*Methods Mol Biol, 231,23*

# Random Mutant Libraries by Megaprimer PCR



***DpnI* digestion**
**(metylated template DNA)**

***denaturation***
**annealing**

***transformation***
***segregation***

*K. Miyazaki (2003)*
*Methods Mol Biol, 231,23*

# Generation of Diversity by Recombination



## *In vitro* Methods

### Random Fragmentation

DNaseI digestion
Stemmer (1998)
*Nature* **370,** 389

### Random priming synthesis

Shao et al. (1998)
*Nucleic Acids Res.* **26,** 681

### Staggered Extension Process (StEP)

Zhao et al. (1998)
Nature Biotechnol. **16,** 258

### RACHITT

**R**andom **C**himeragenesis on **T**ransient **T**emplates

Coco et al. (2001)
*Nat. Biotechnol.***19,** 354

## *In vivo* Gene Recombination

### Site specific recombination

*cre - lox*

lambda

### Homologous recombination

E.coli
Phage Lambda
*Saccharomyces cerevisiae*

### Recombination of Non-Homologous sequences

ITCHY
SCRATCHY
SHIPREC

# Generation of Diversity by Recombination



## *In vitro* methods

### Random fragmentation

- DNase I digestion;
  Stemmer (1998) *Nature* **370,** 389

### Random priming synthesis

Shao *et al.* (1998) *Nucleic Acids Res.* **26,** 681

### Staggered Extension Process (StEP)

Zhao *et al.* (1998) Nature Biotechnol. **16,** 258

### **RACHITT** (**R**andom **C**himeragenesis on **T**ransient **T**emplates)

Coco *et al.* (2001) *Nat. Biotechnol.***19,** 354

## *In vivo* gene recombination

### Site-specific recombination

Cre-lox; Lambda

### Homologous recombination

*E.coli*, Lambda phage, *Saccharomyces cerevisiae*

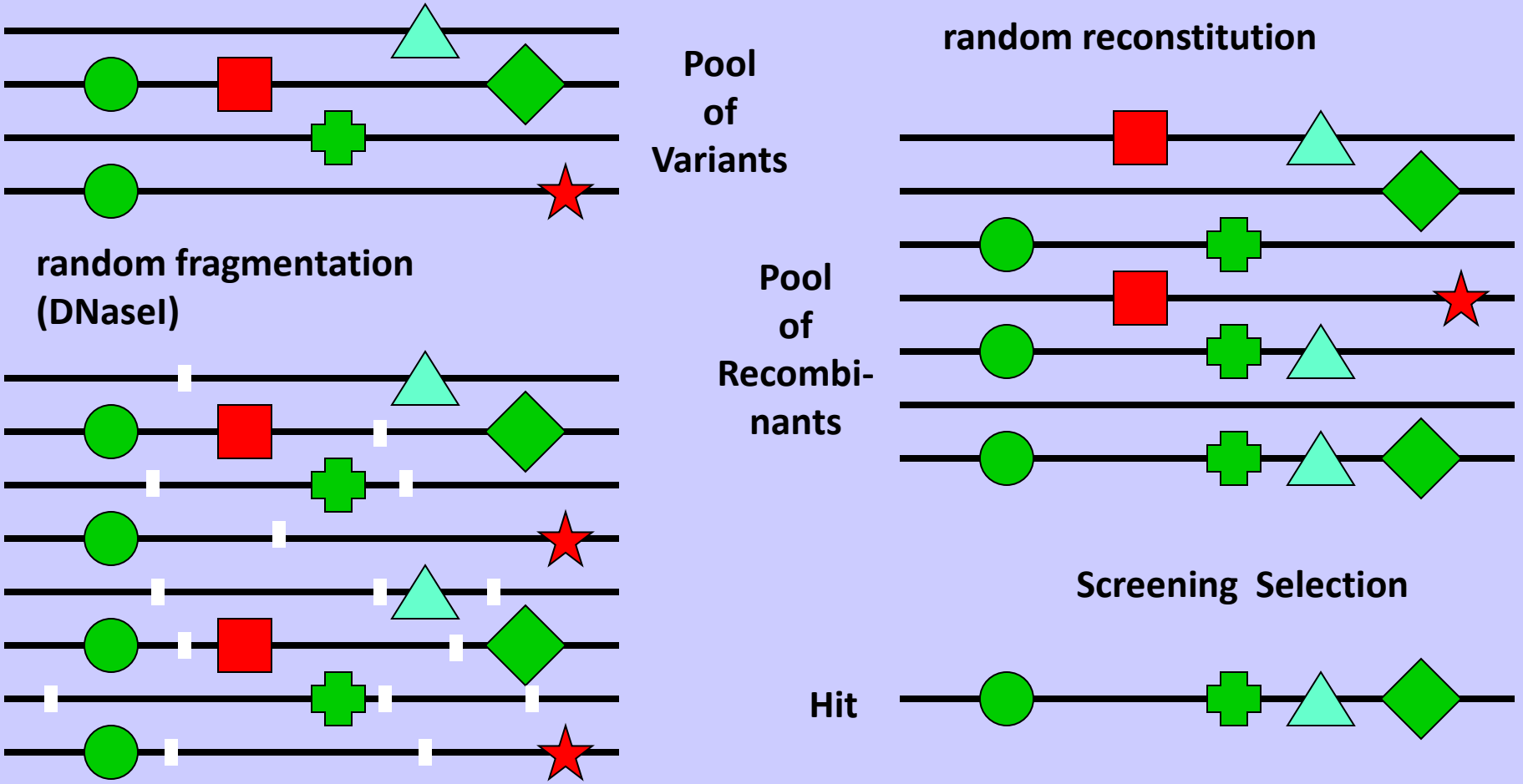
### Recombination of non-homologous sequences

ITCHY, SCRATCHY, SHIPREC

# Molecular Evolution of Enzymes

## In vitro Recombination of Sequences

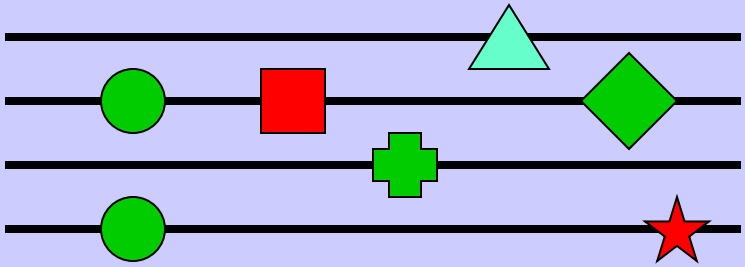
"Gene Shuffling" "Family Shuffling"



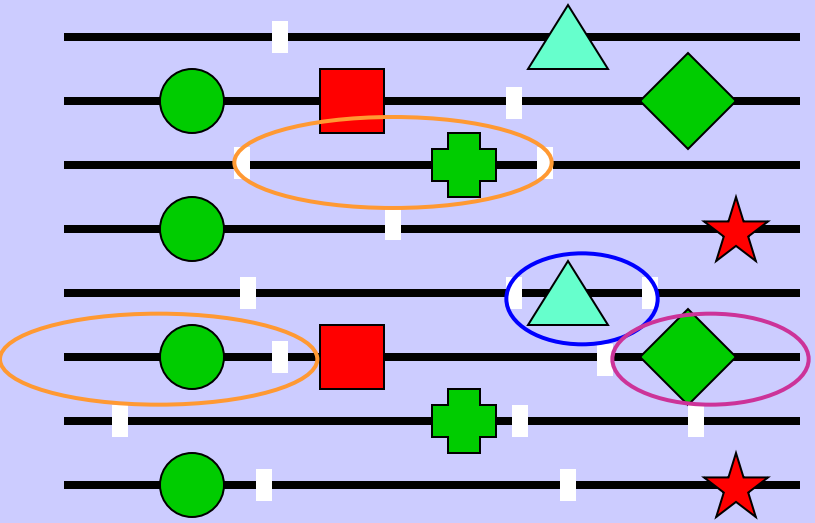
# Molecular Evolution of Enzymes

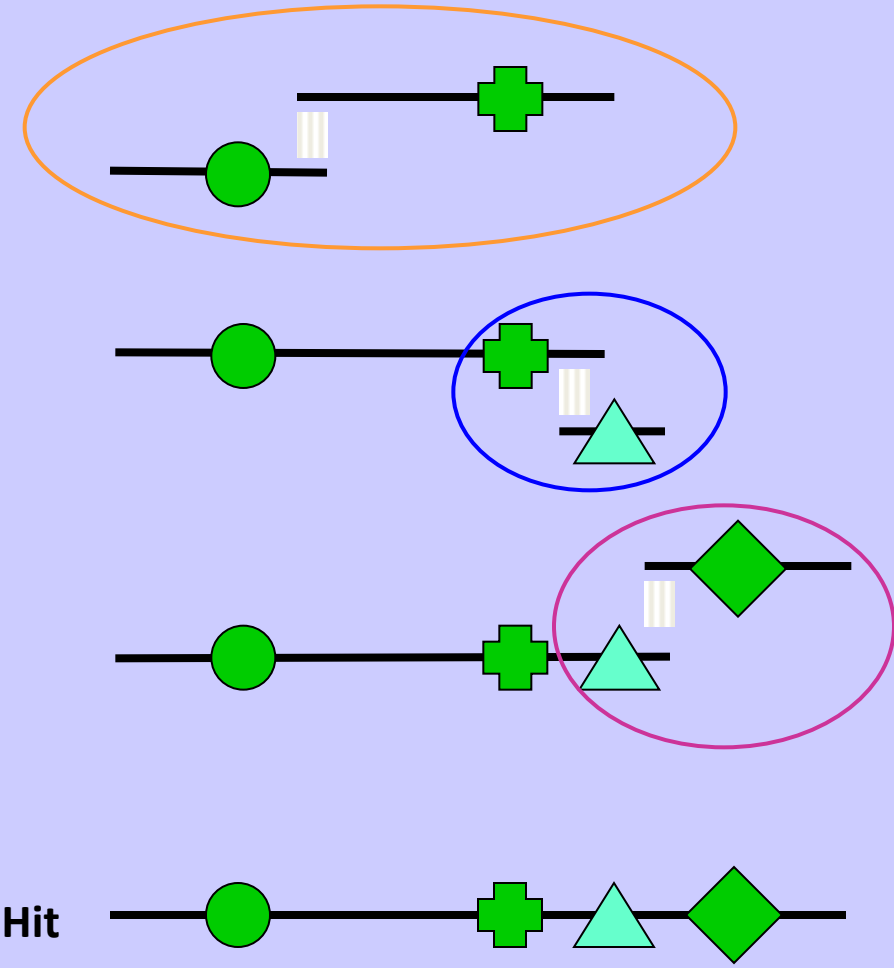
## In vitro Recombination of Sequences

"Gene Shuffling" "Family Shuffling"



random fragmentation
(DNaseI)



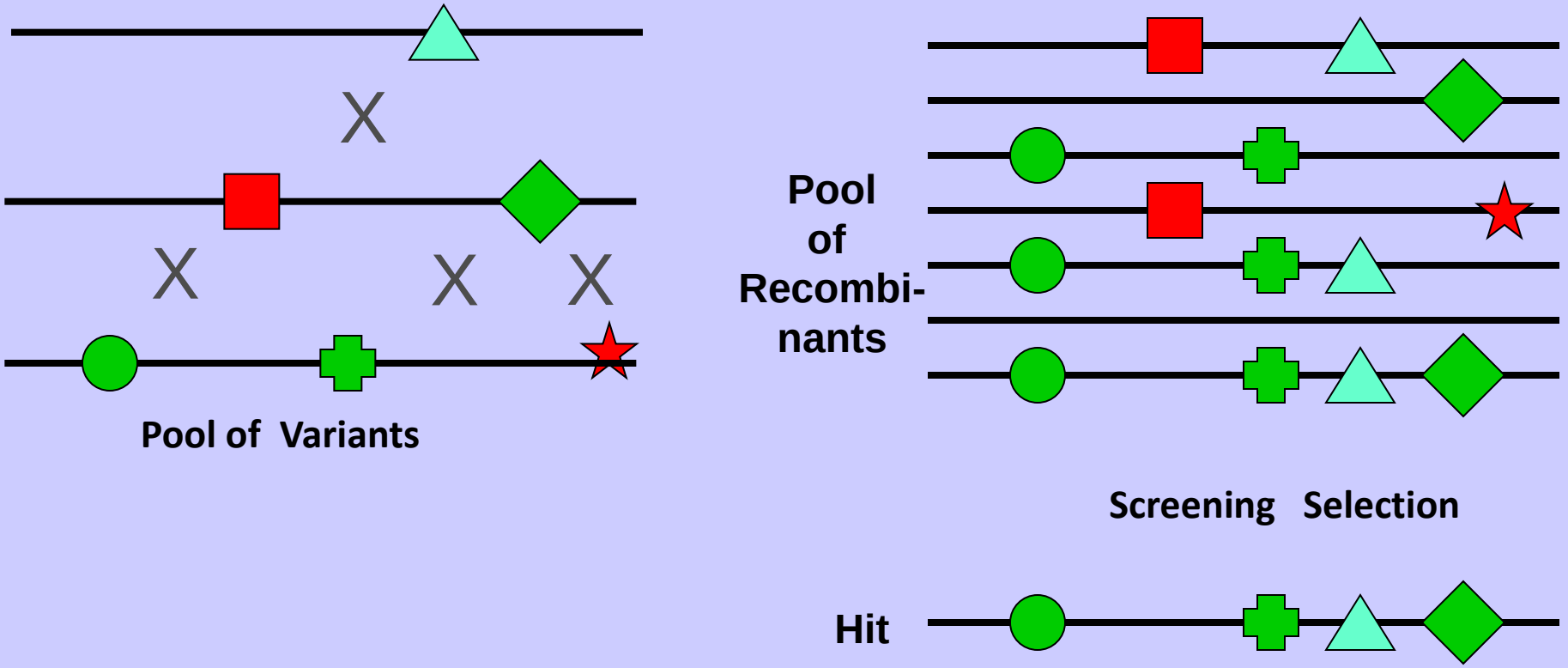


# Molecular Evolution of Enzymes

## In vivo Recombination of Sequences

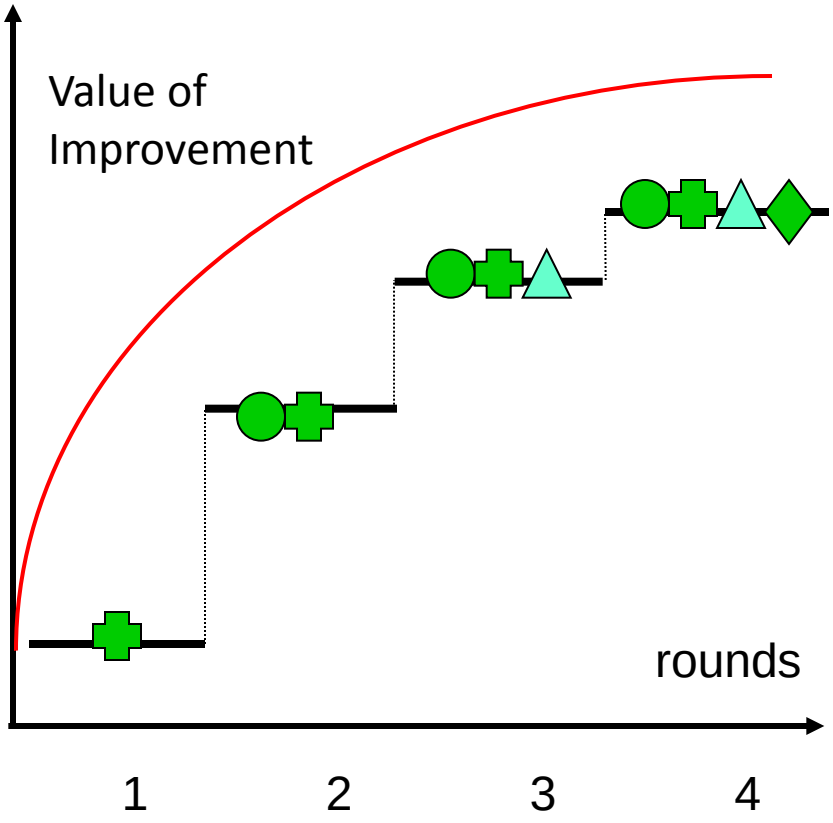
"Gene Shuffling" "**Family Shuffling**"

**random recombination**

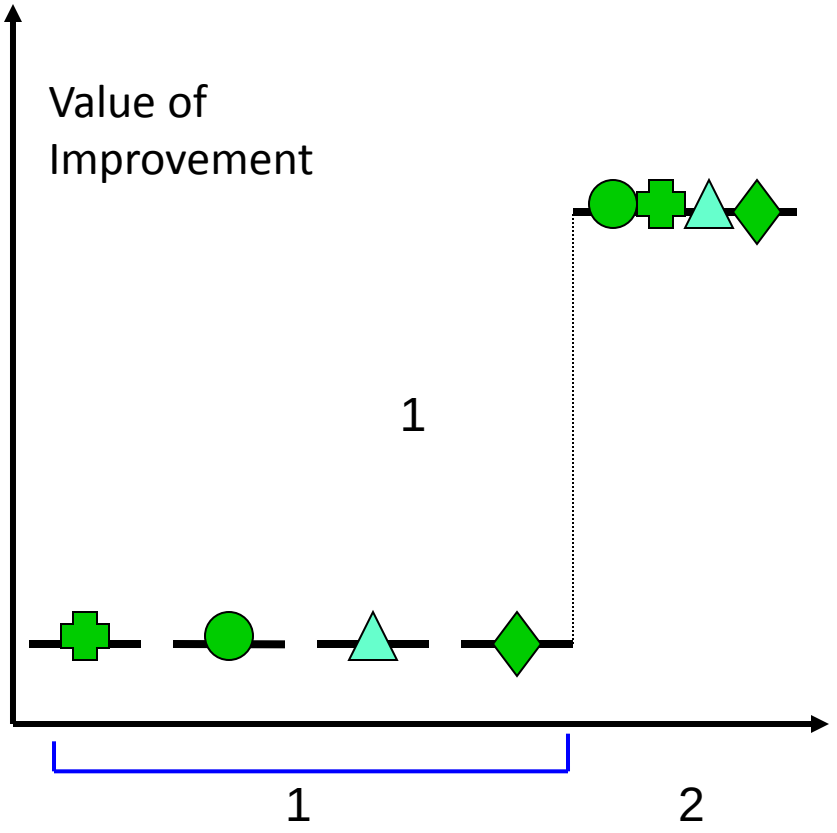


# Molecular Evolution of Enzymes

## Rounds of Random Mutagenesis

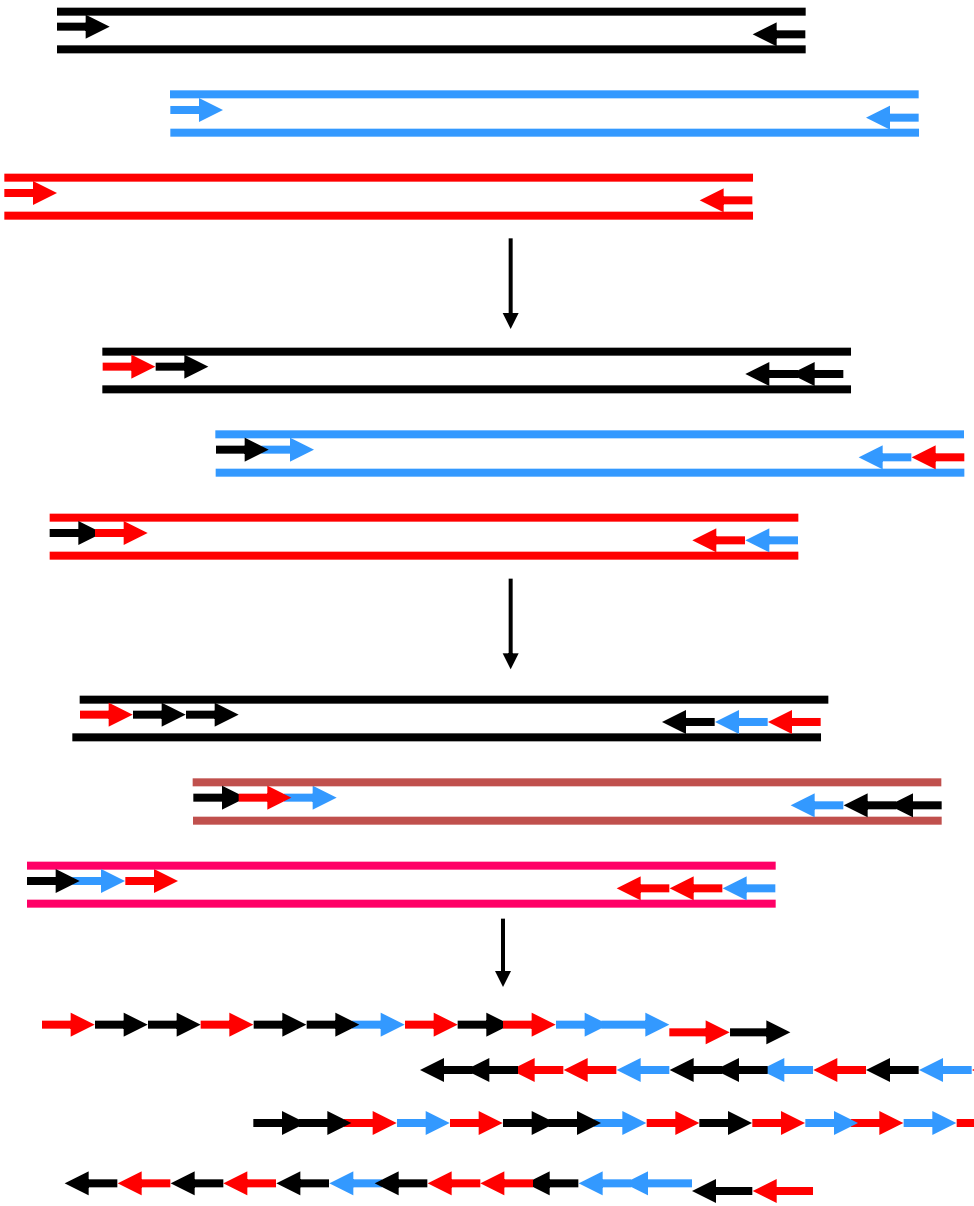



## Recombination of Sequences



30.4.15

# Random Recombination Libraries by StEP



***round 1***
***- denaturation***
**- primer annealing**
**- extension**

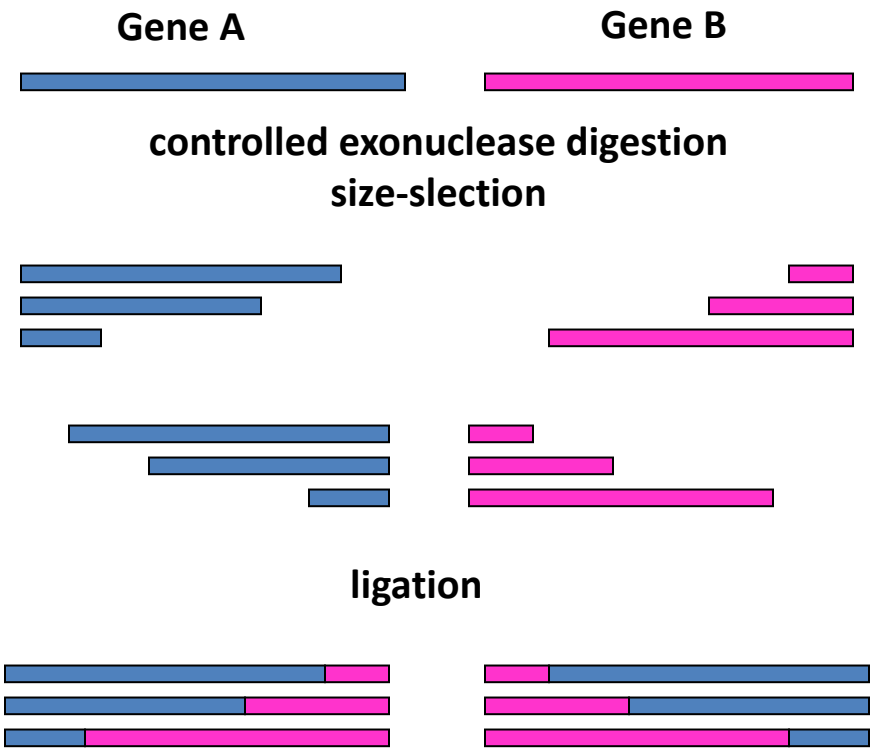
***round n***
***- denaturation***
**- primer annealing**
**- extension**

**Library of mosaic genes**

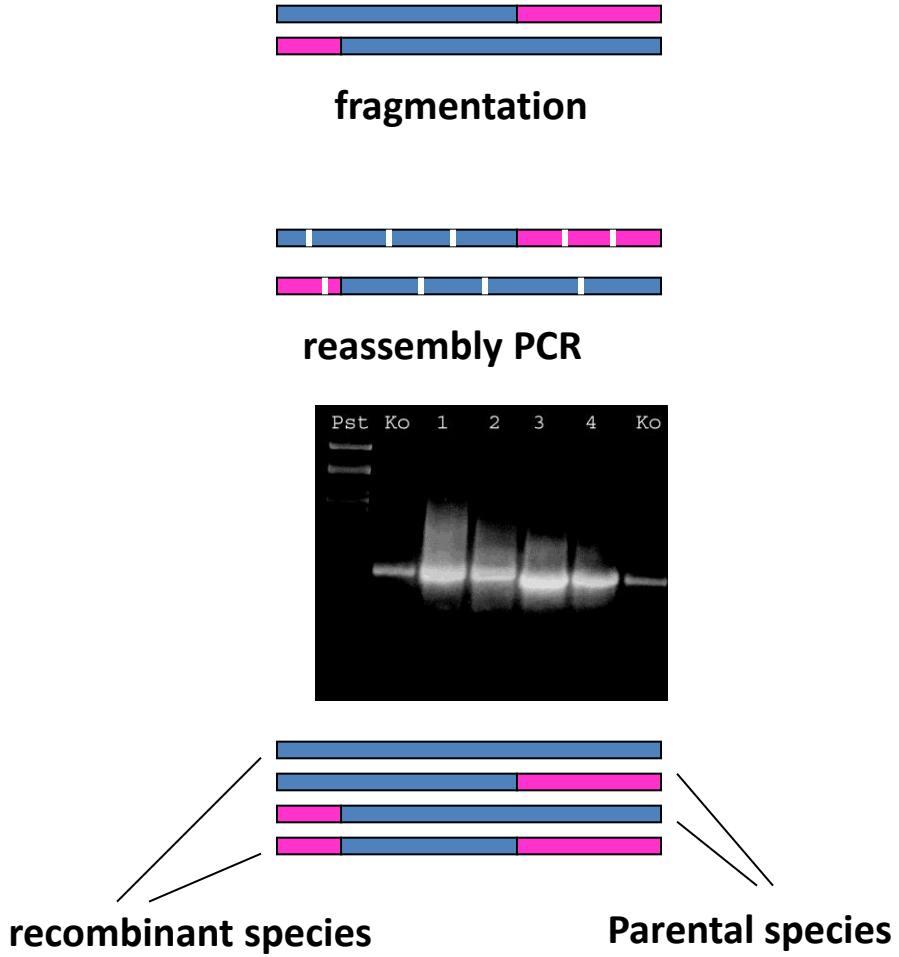
*Aguinaldo & Arnold (2003)*
*Methods Mol Biol, **231**,105*

# Recombination of non-homologous sequences

**Step 1:**
**random truncation - fusions**



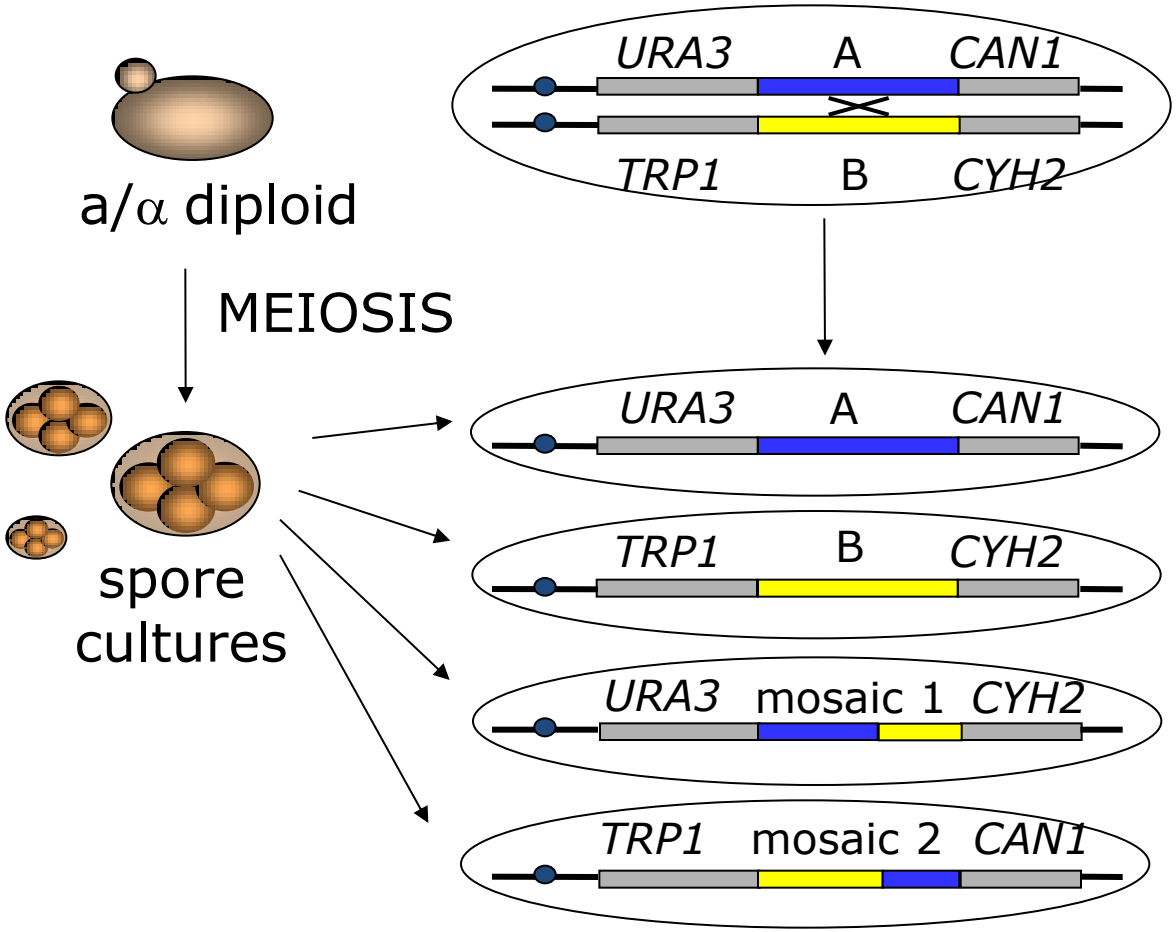

**Step 2:**
**Shuffling of fused variants**




Pool of random point fusions

*B.Brunner, 2001*

# Overview of the yeast shuffling strategy




**Meiosis :**

high levels of genome-wide recombination

***MSH2*** **:**
**key player in homeologous recombination, mismatch repair and mutagenesis**

## Screening - Selection

First law of directed evolution:

**You get what you screen for**

Frances Arnold

# Screening - Selection

## Screening: Individual Analysis of clones

### Essentials:

- high throughput - HTP
- simple and robust
- good discriminatory capacity
- accessible to robot handling
- application of process-near conditions (e.g. organic solvents)
- allows work with desired substrate → no surrogates

## Selection: Growth advantage

### Prerequisite:

**Bio-compatibility**

# Screening – Selection: Problems to consider

- **Uniform growth of individual clones**
  - Substrate supply
  - Mass transfer (oxygen supply, $CO_2$ emmission
  - heat transfer (e.g. position on plate/shaker)
  - inoculation
  - cross contamination
- **Homogeneous expression levels**
  - Host system
  - Vector copy number
  - Induction conditions
  - Functional expression (e.g. folding, post-translational modifications
- **Equal access/release of reactants to/from enzyme**
  - membrane/cell wall transfer
  - cell disruption

# Screening Systems → hosts

## New Hosts for Enzyme Screening

### Bacterial Hosts

| | | |
|---|---|---|
| *Bacillus* strains | → | secretory enzymes |
| *Streptomyces* sp. | → | expression background |

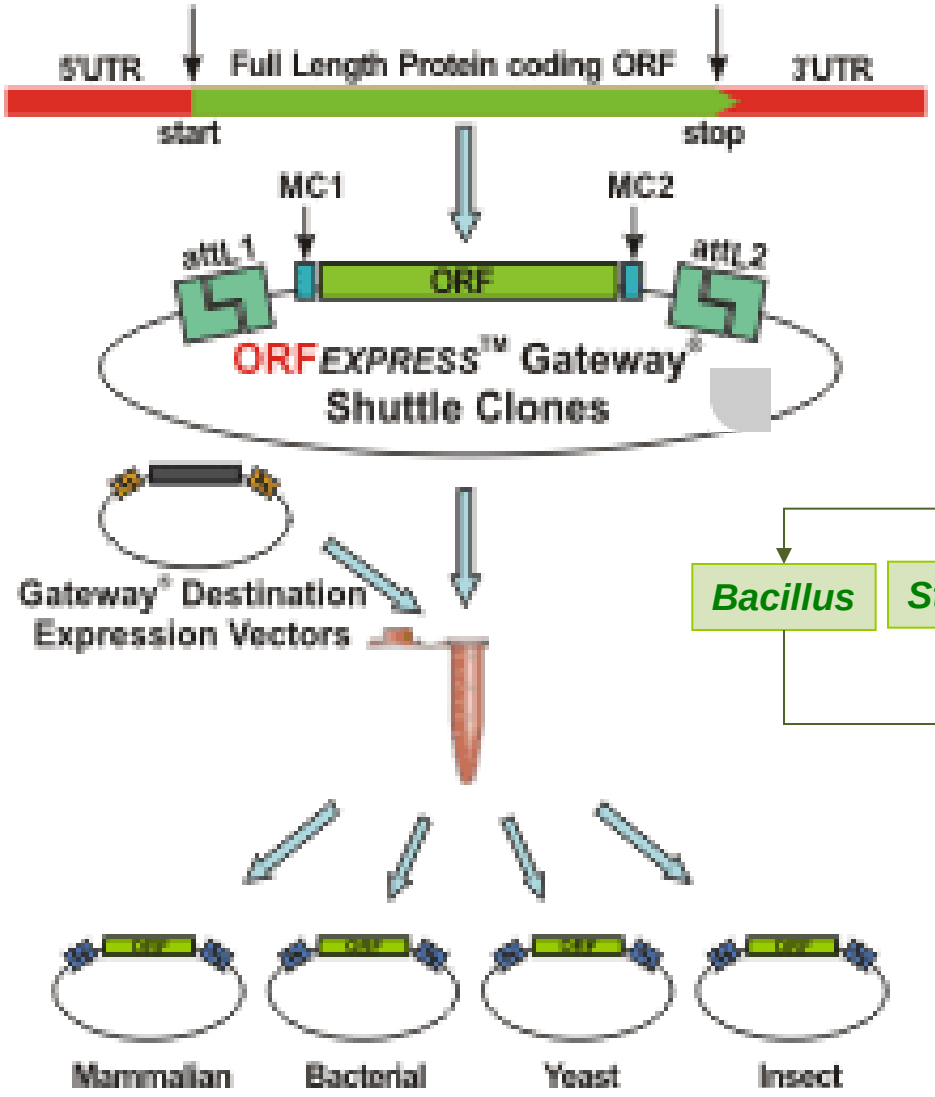
### Fungal Hosts

| | | |
|---|---|---|
| *Pichia pastoris* | → | enzymes of eukaryotic origin |

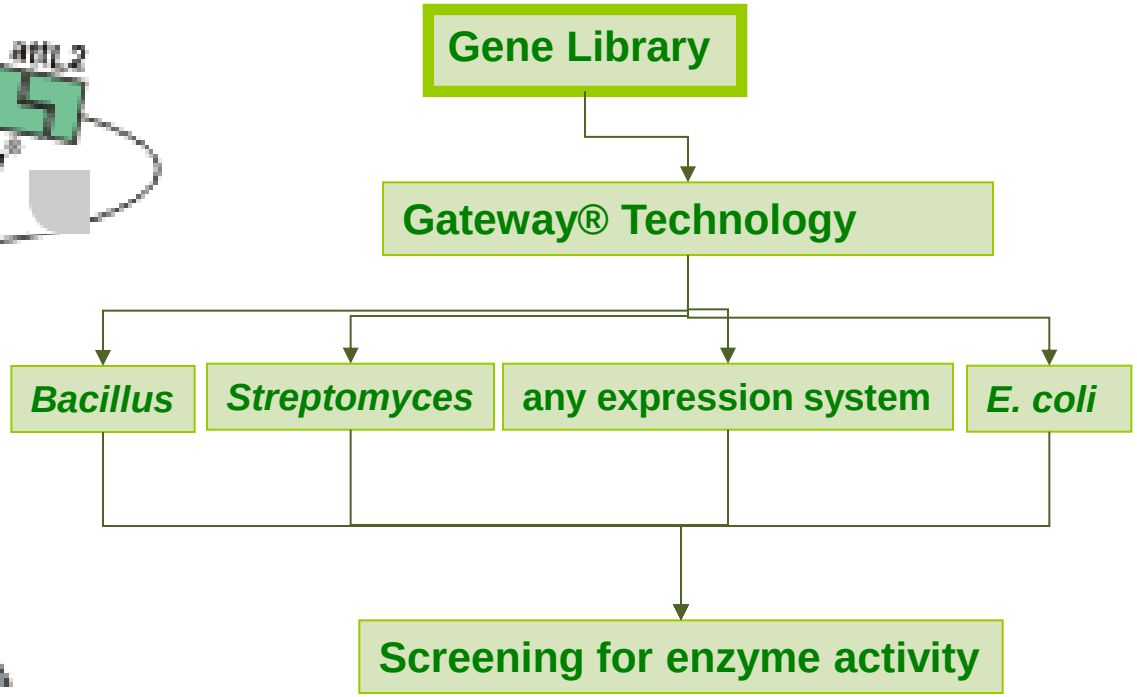
### New Library Concepts

Gateway Technology
in vivo Transfer Systems
Genome integration

# Screening Systems → hosts





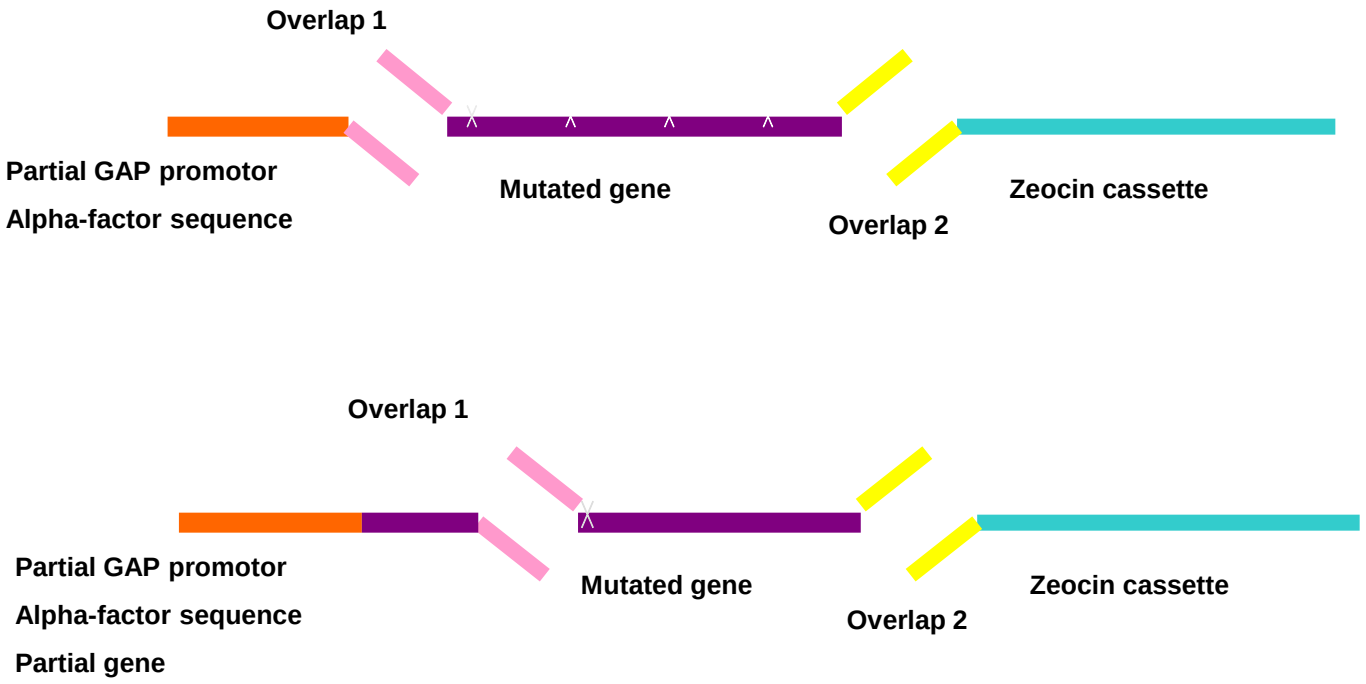




**Gateway® Technology .**

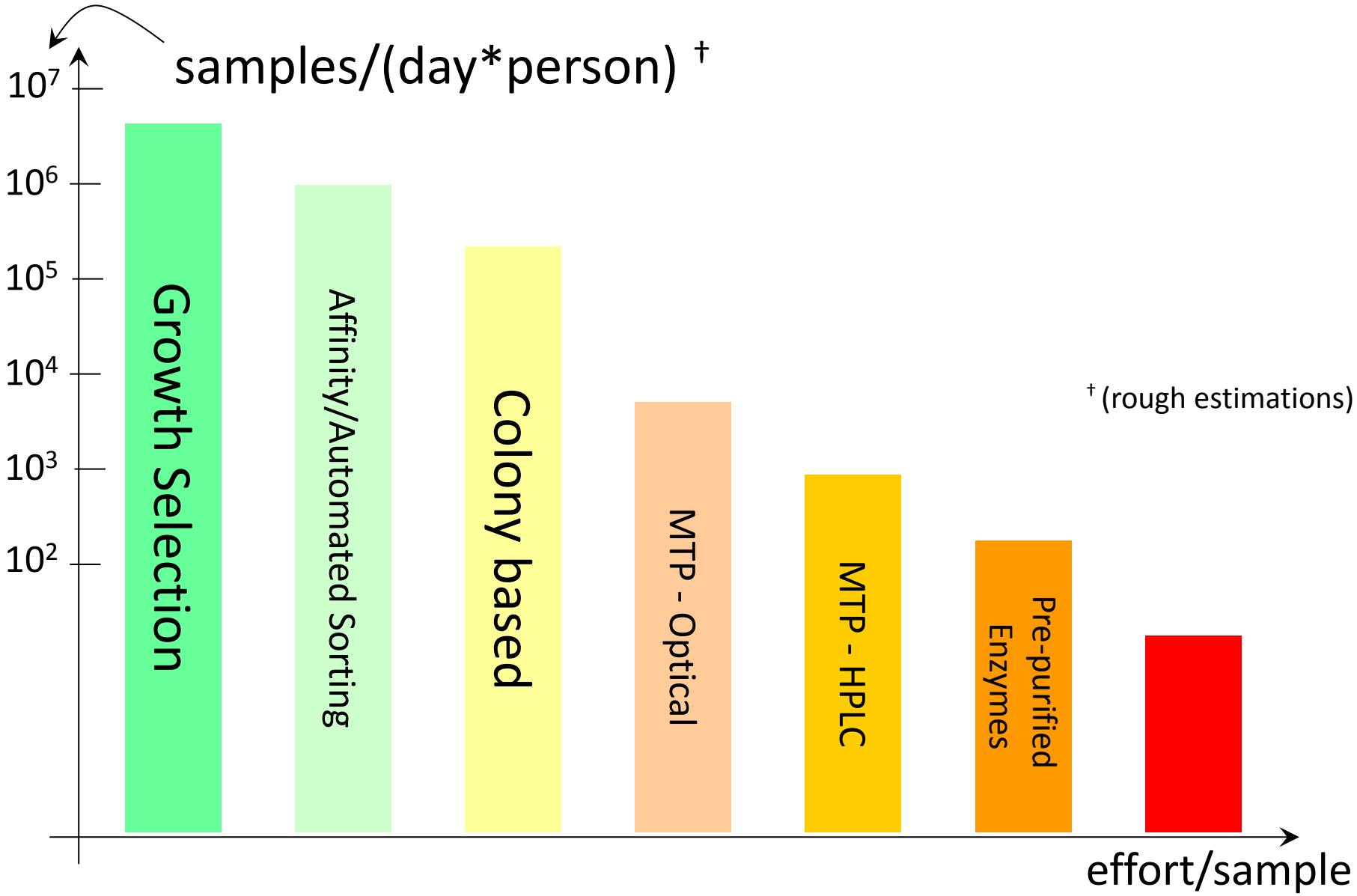
# New Strategy for Library Generation

## Random & Site Directed mutagenesis

### Directed Evolution in *Pichia pastoris*

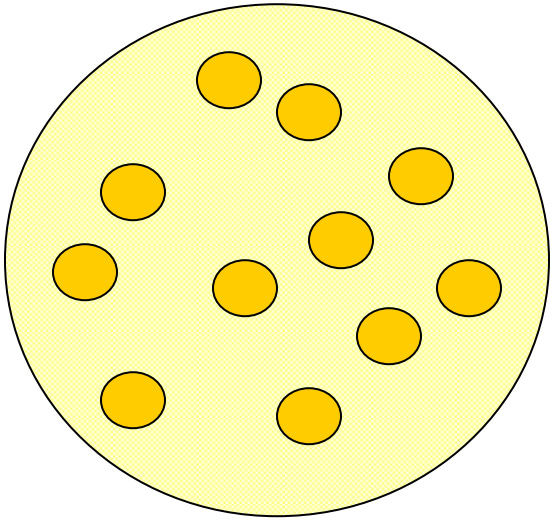


# High Throughput Screening -Detection of Enzymatic Activity



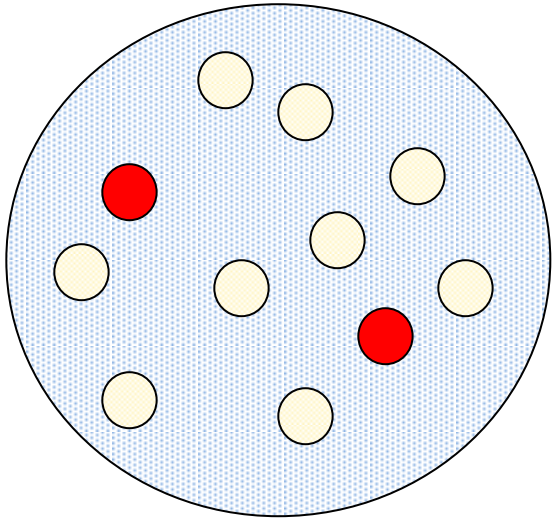
# Screening by filter assays

## Colonies on agar plates



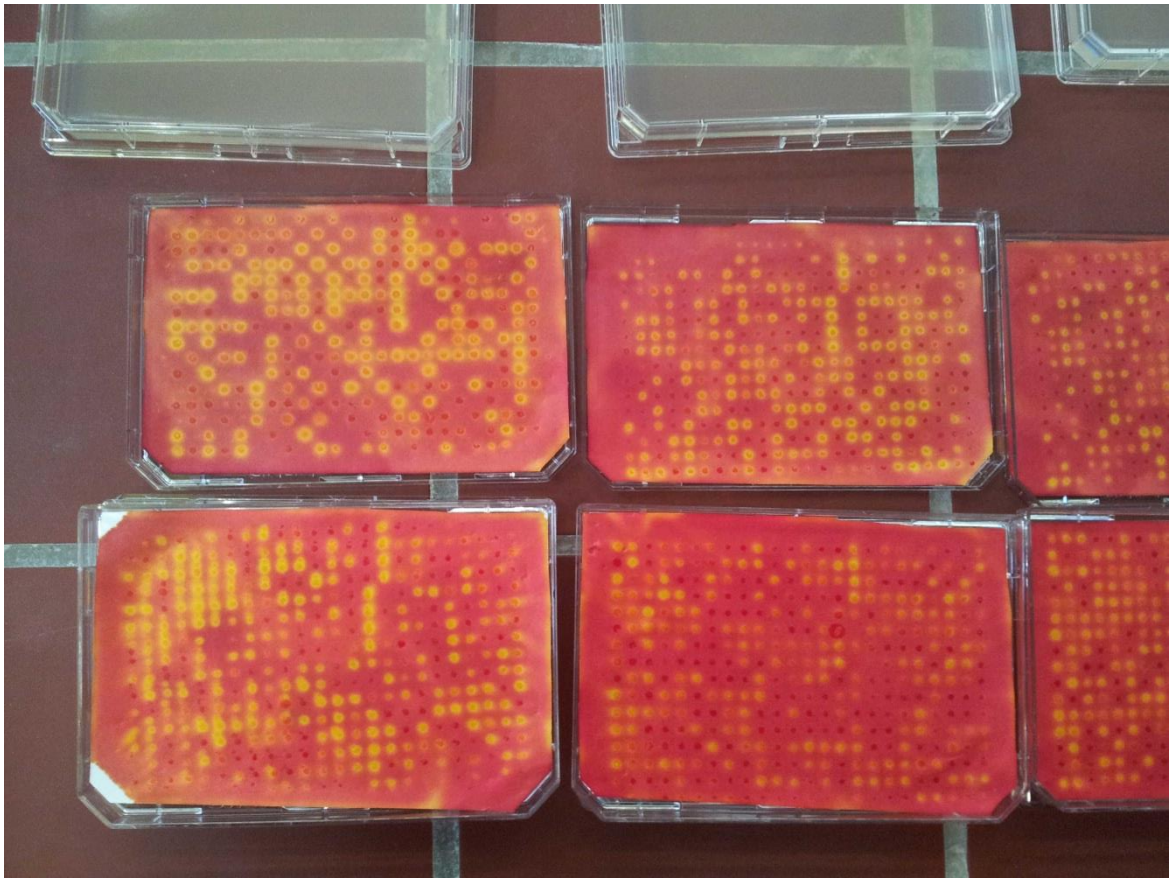
## Transfer to filter

## Substrate – Reaction Detection



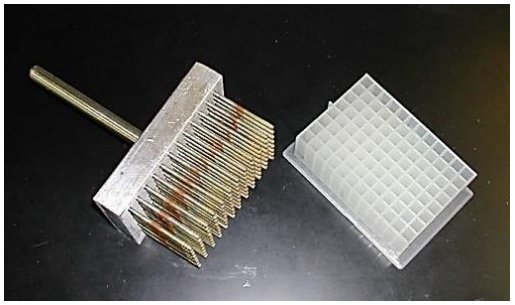
# Screening by filter assays

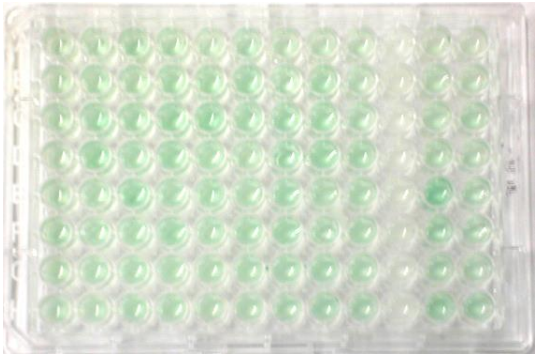
Screening based on detection of pH shift
Example: Esterases



# Screening - Selection

## Cultivation in Liquid Culture



**Deepwell plates**

**Shake flasks**

**Lab fermenters**

**Detection**

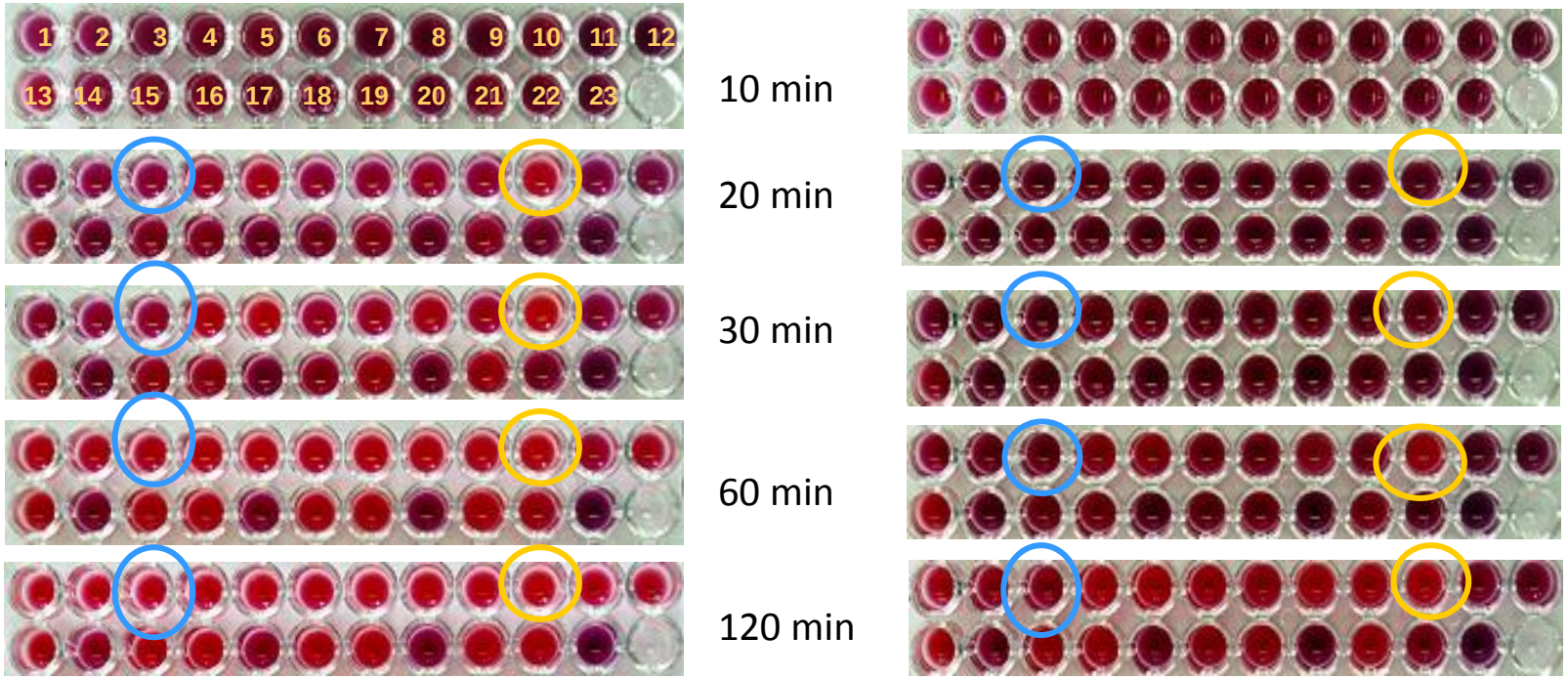
**Microplate assays – Photometric Fluorometric**
**HTP chromatography (e.g. HPLC)**
**HTP MS methods**
**HTP NMR methods**

# Screening systems

## Microplate Assays

### *rac*-linalyl acetate

### (*R*)-linalyl acetate




**22: wt enzyme**

**23: blank lysate**

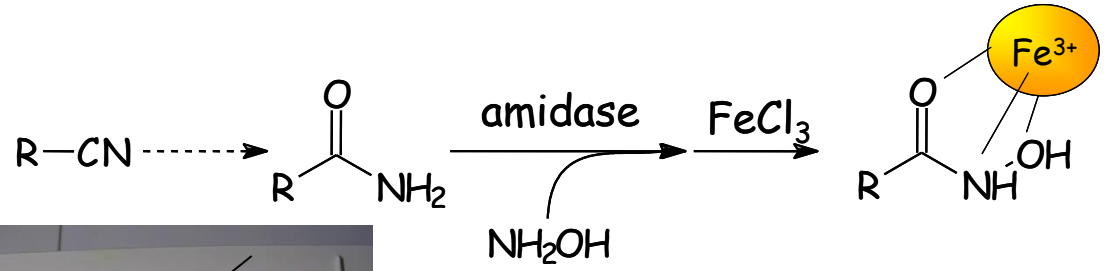
5.5.15

# New Nitrile Hydratases

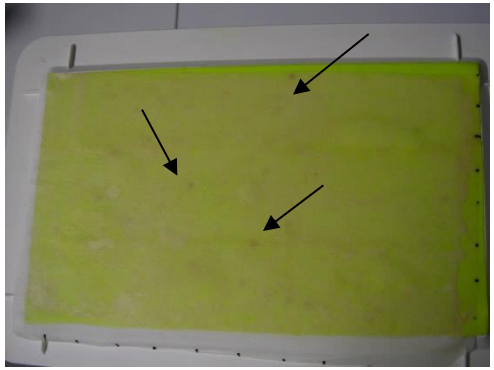
2.1

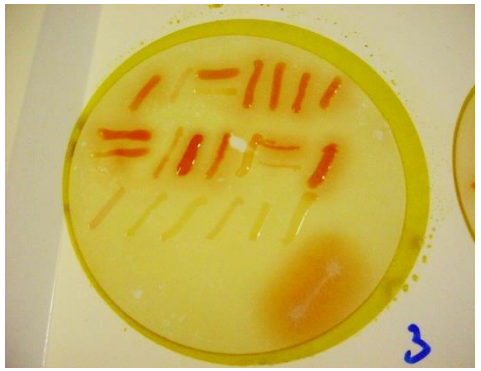
## Screening assays

*In vitro* screening assay

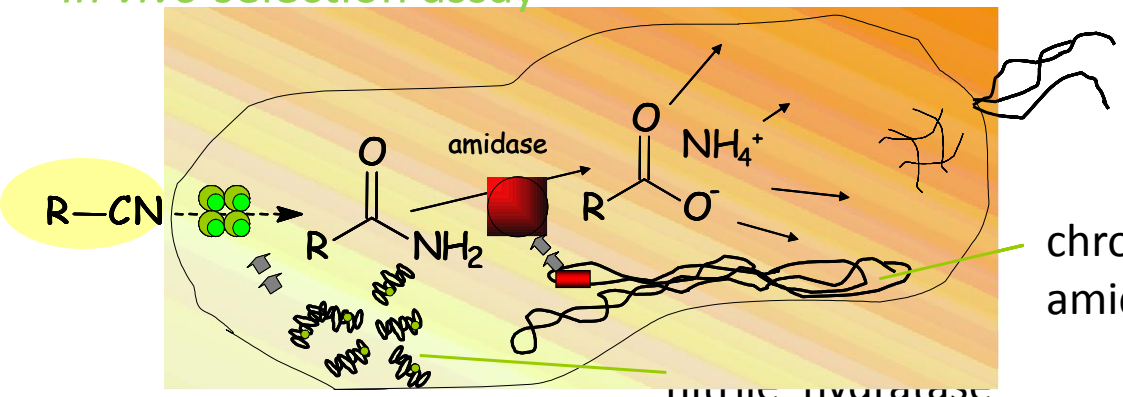



hydroxamic acid





*In vivo* selection assay

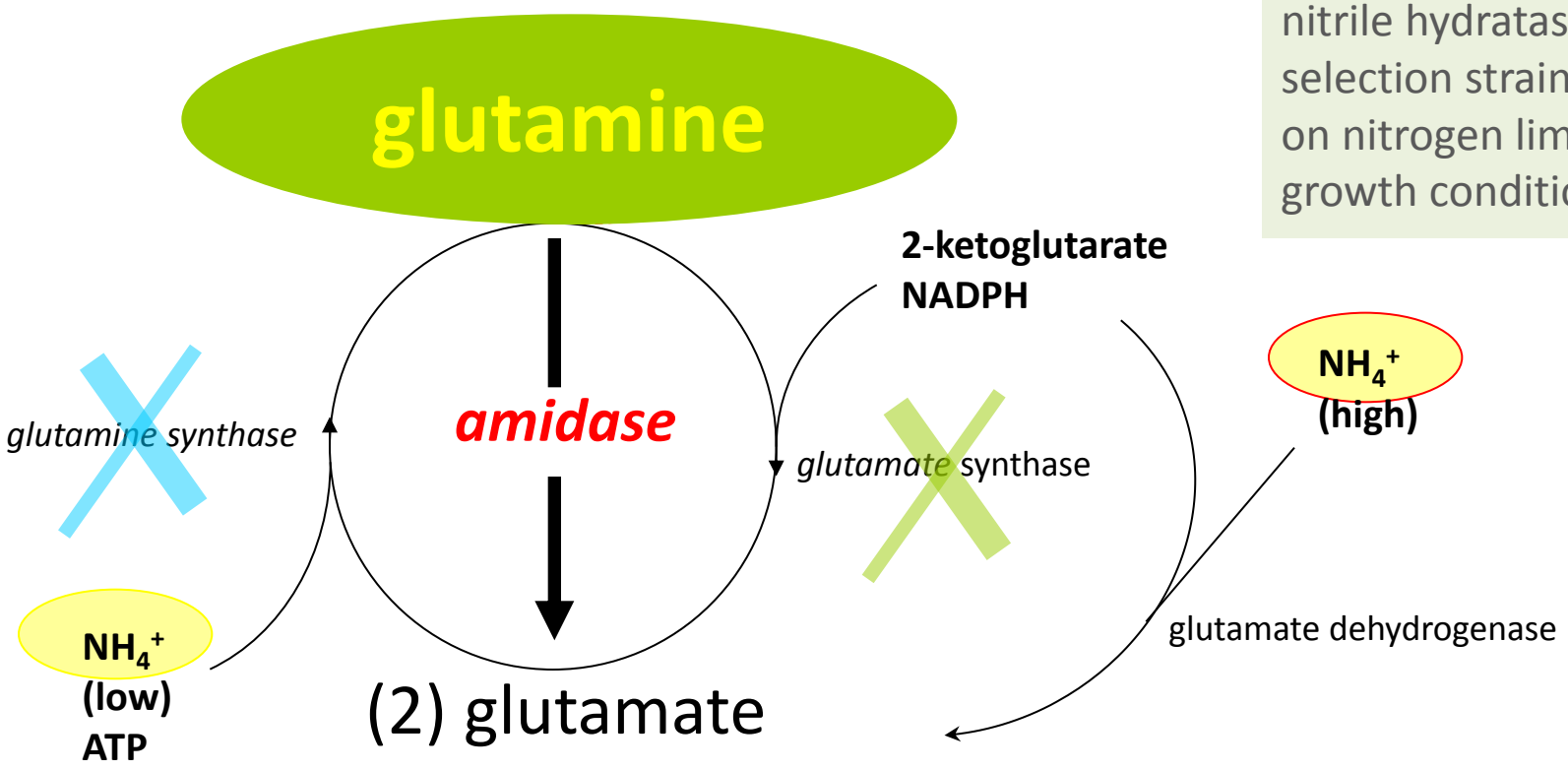



chromosome with amidase integration

nitrile hydratase expression plasmids

# Selection Systems

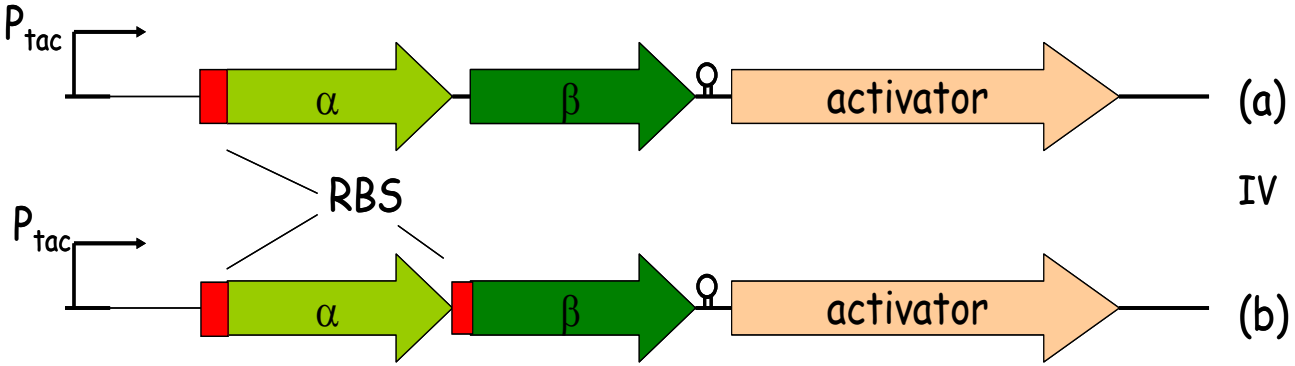
Development of a nitrile hydratase selection strain based on nitrogen limited growth conditions

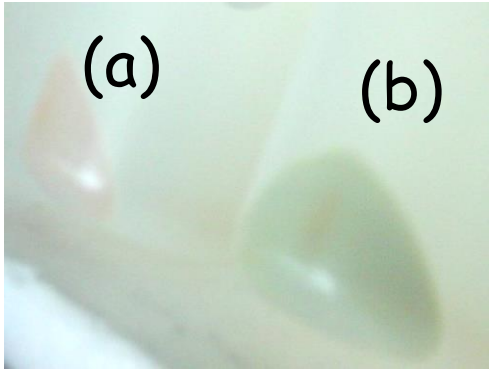



nitrile —*nitrile hydratase*→ amide —*amidase*→ acid + $NH_4^+$

# New Nitrile Hydratases

2.1

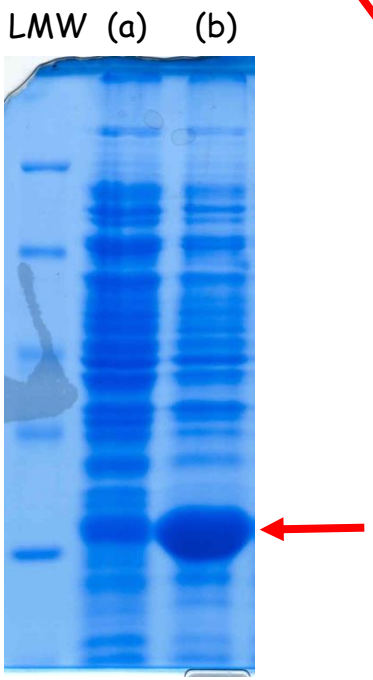



**Improvement of the nitrile hydratase expression level by re-design of the expression cassette**

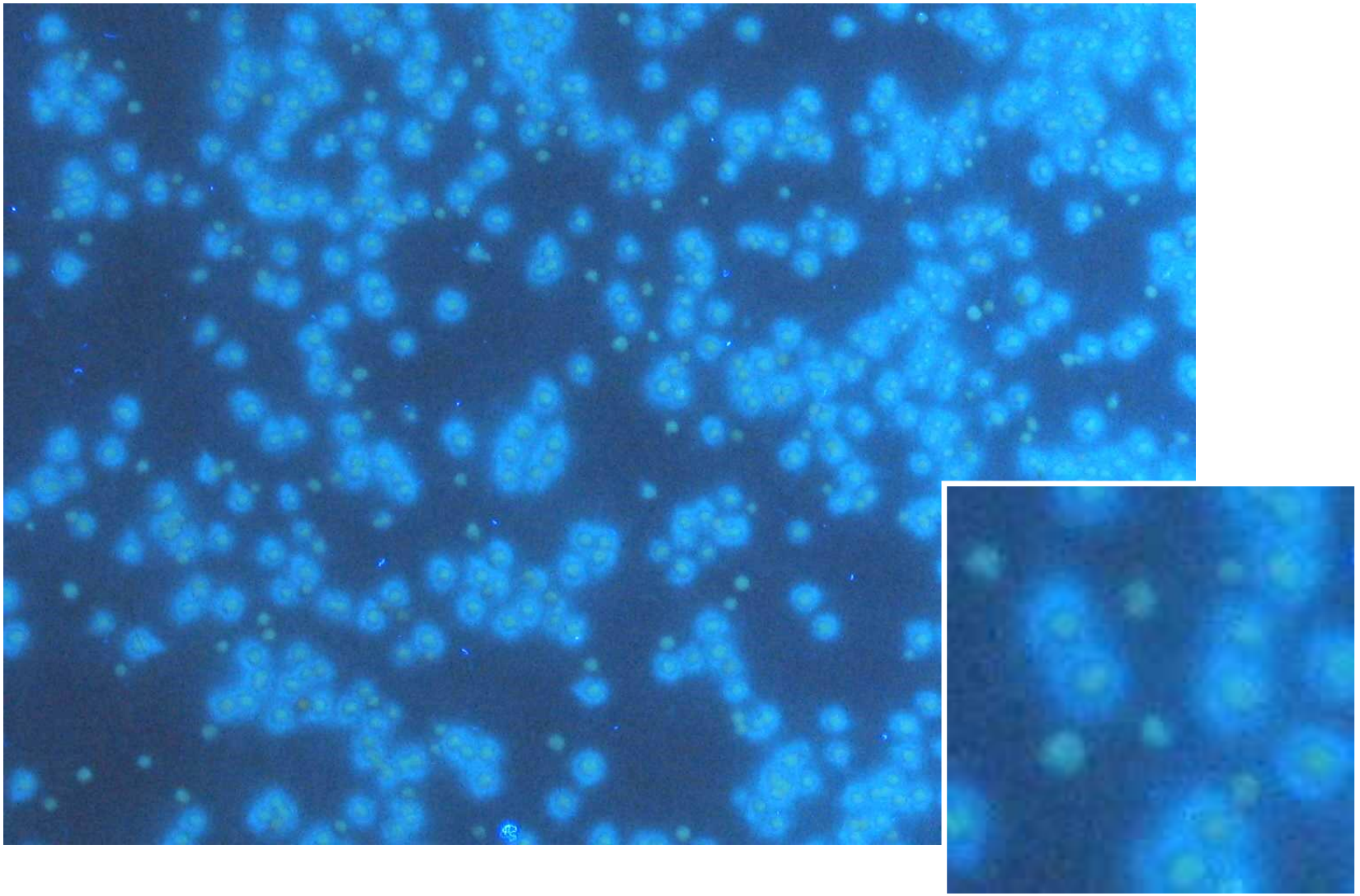
Introduction of a second copy of the optimized ribosome binding site upstream of the beta subunit

→ **dramatic increase** in soluble (and active) nitrile hydratase formation.
→ green appearance of the pellet of cells producing high amounts of the iron containing enzyme.



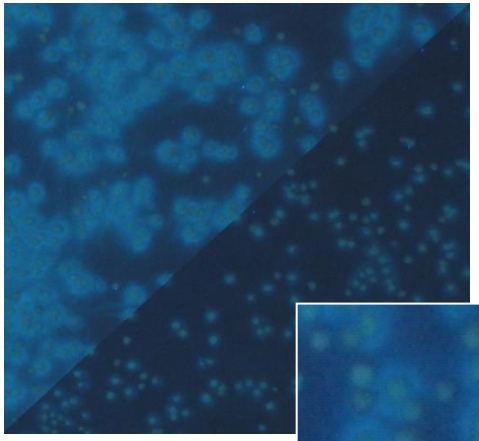

**→ Industrial Application**

# NADH Fluorescence Coupled Assay

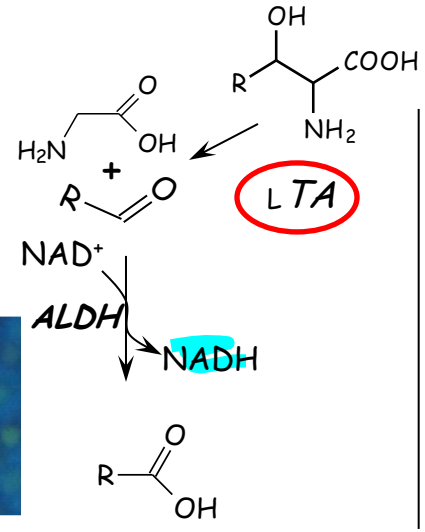


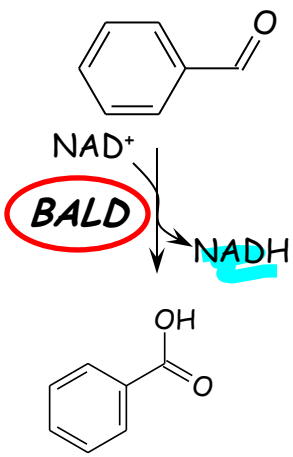
Mixture of *E. coli* colonies with and w/o threonine aldolase activity

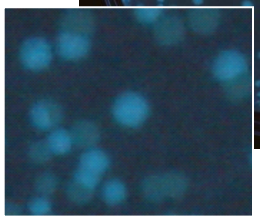
# NADH – A Versatile Reporter

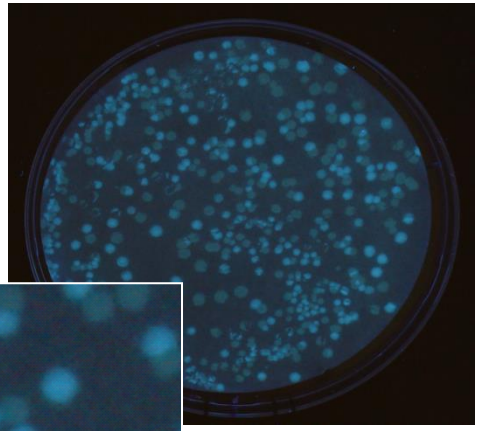


L-threonine aldolase



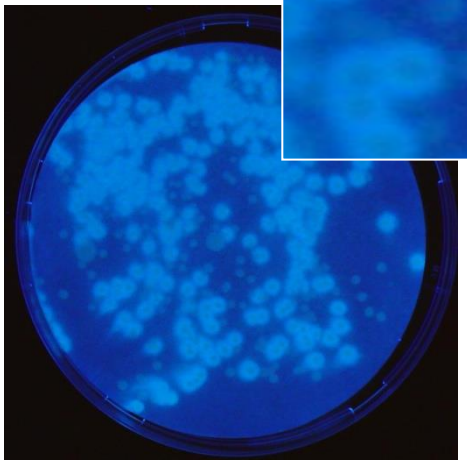



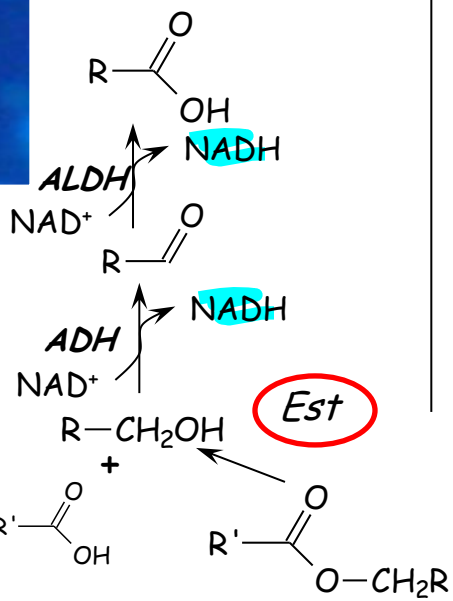


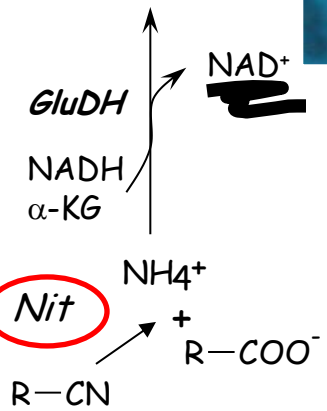
(benzaldehyde) dehydrogenase

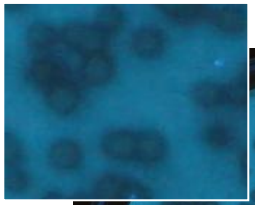
esterase







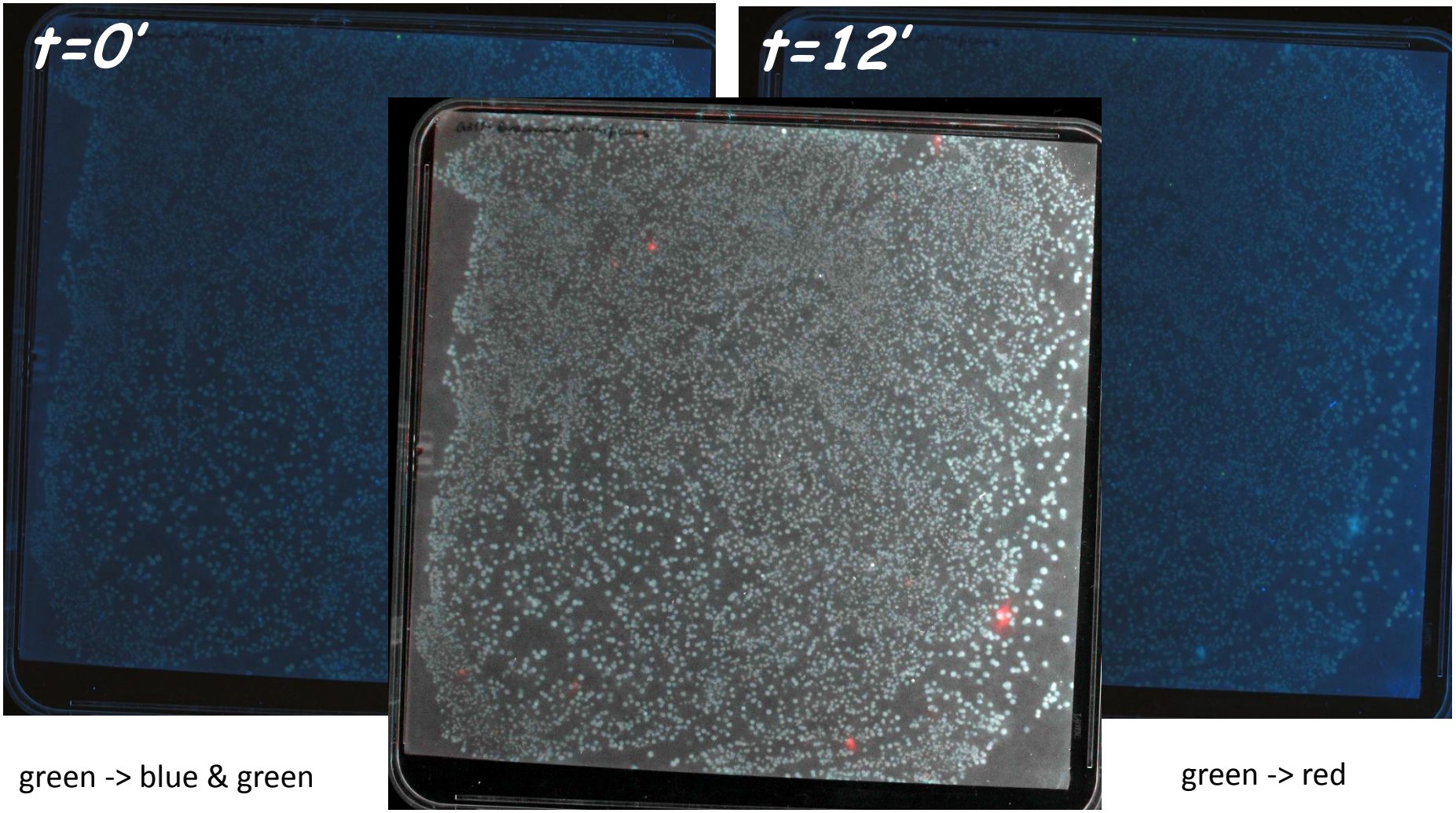



nitrilase



Reisinger *et al. J. Mol. Cat. B,* **2006**, 39, 149-155

# Data evaluation

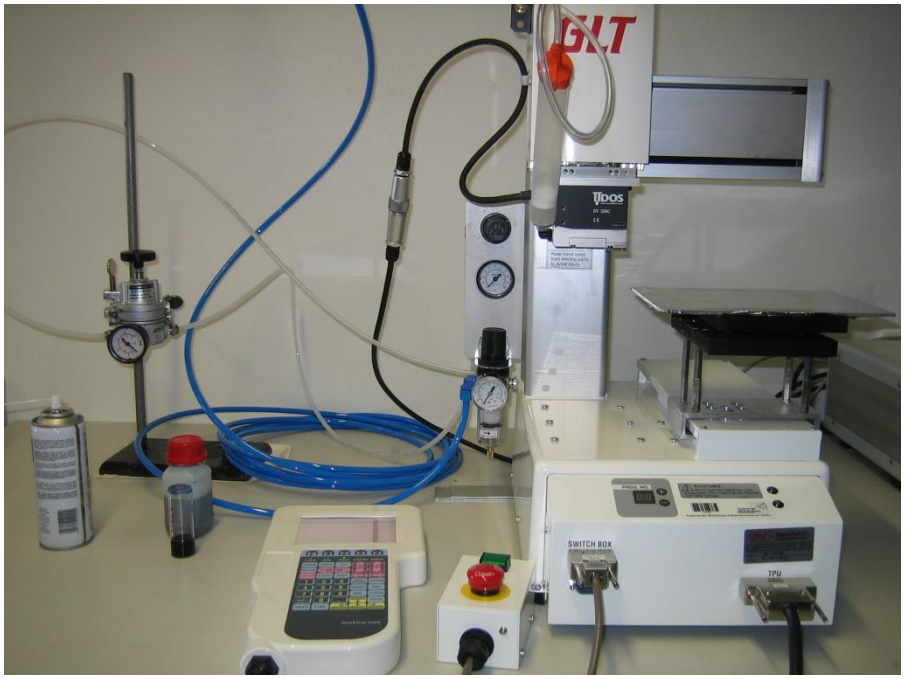


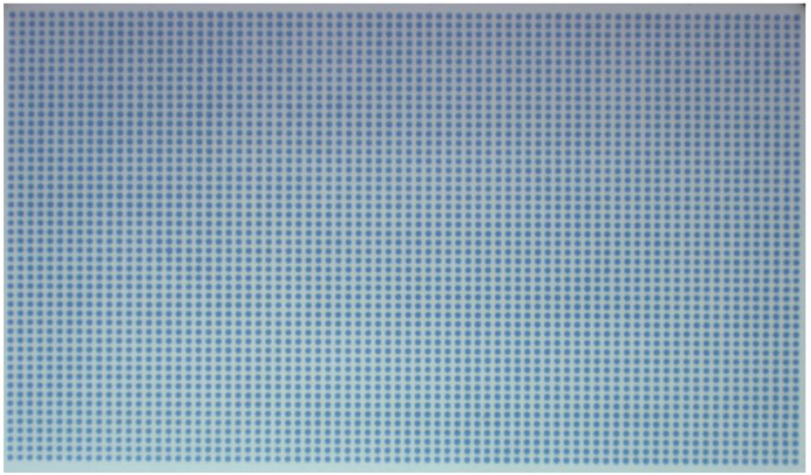

green -> blue & green

green -> red

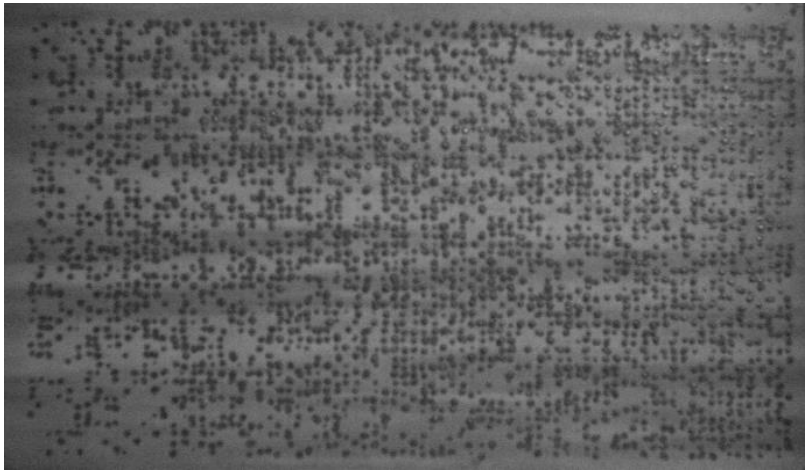
*1.Rasband, W.S., **ImageJ**, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2006.*

# Micro-colony Arrays



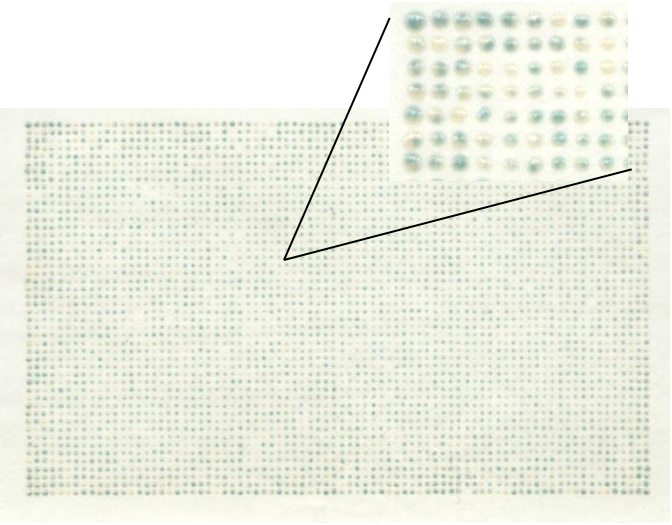
Ordered Arrays

# Micro-colony Array Screening Platform

## High throughput detection of enzyme activity

**Generation of a high density ordered colony array**



6000 colonies on filter (microplate formate)

Direct spotting from library pool – no colony picking

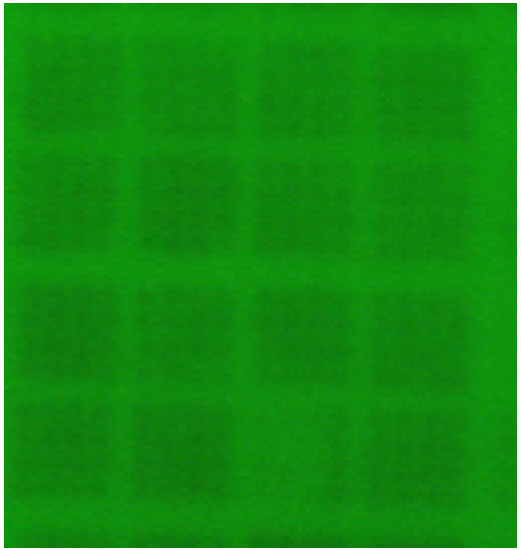
Replicas of filters

Pre-treatment of arrays possible (e.g. solvents, T , pH, etc.)

Enzyme reaction

Detection system (e.g. chemosensor)

**Simultaneously monitoring CCD camera**



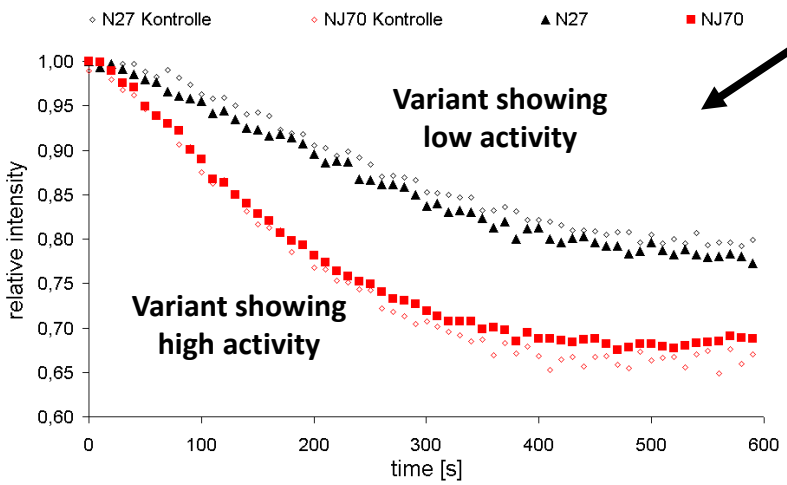
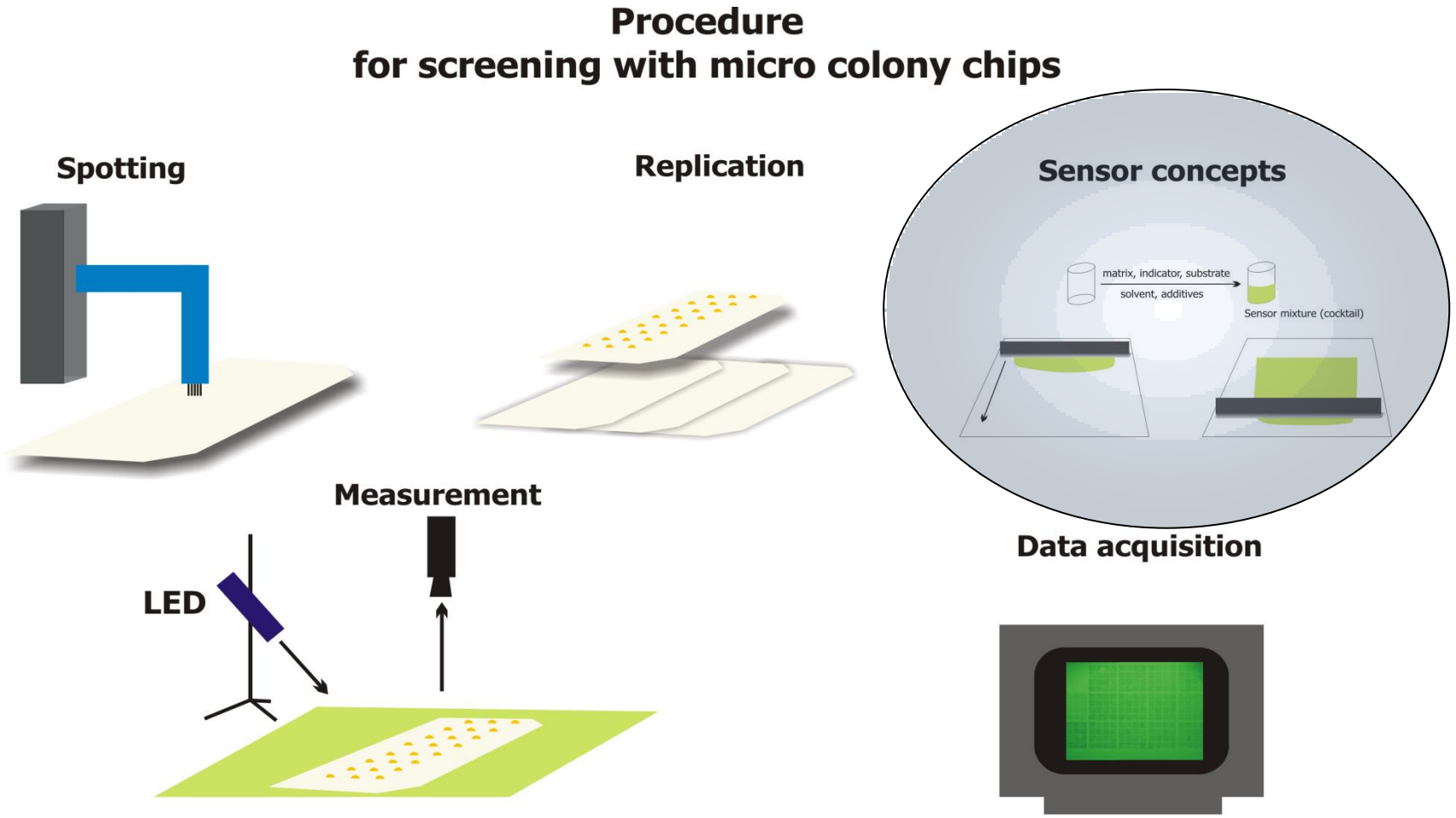
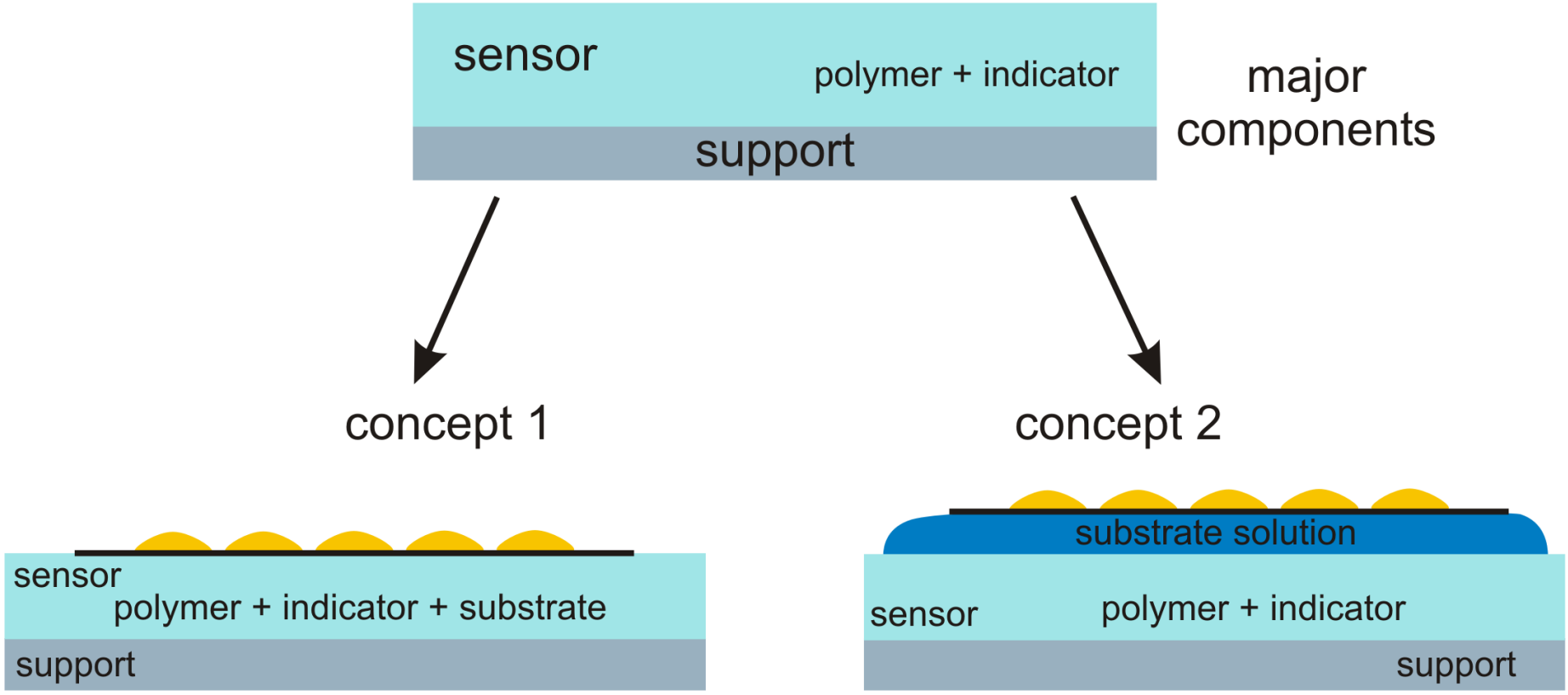



**Image analysis & automated hit detection**

# Micro-colony Array Screening Platform

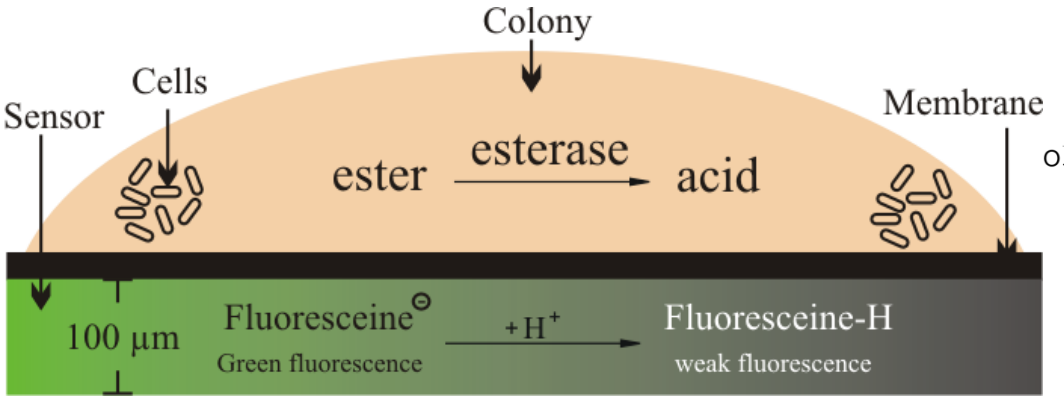


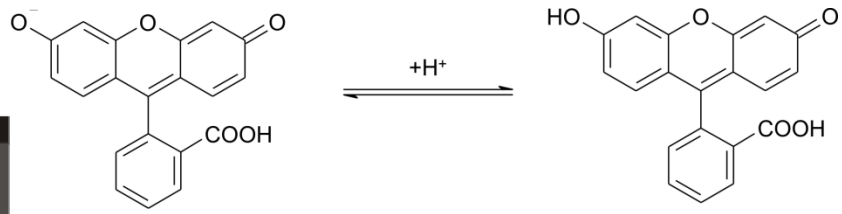
# Sensor design

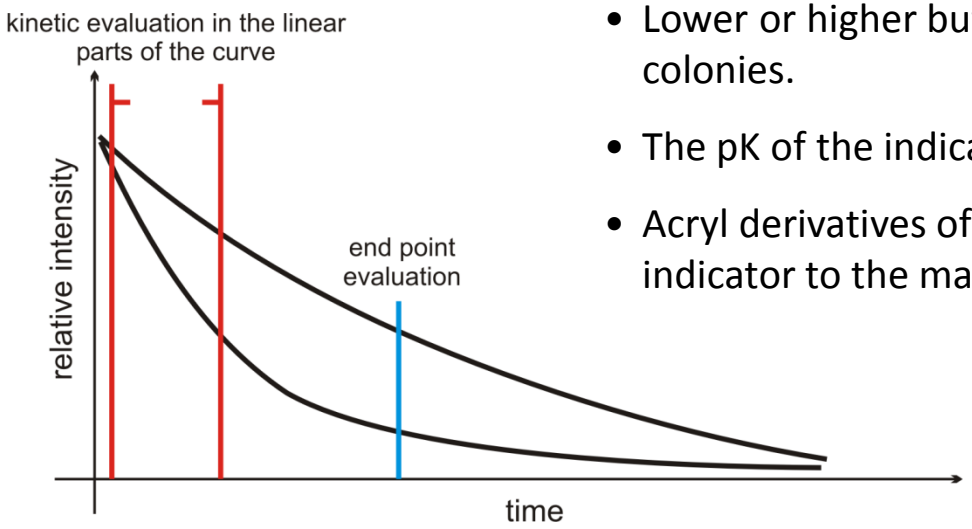


# Sensor design: pH-Sensor

## Esterase screening



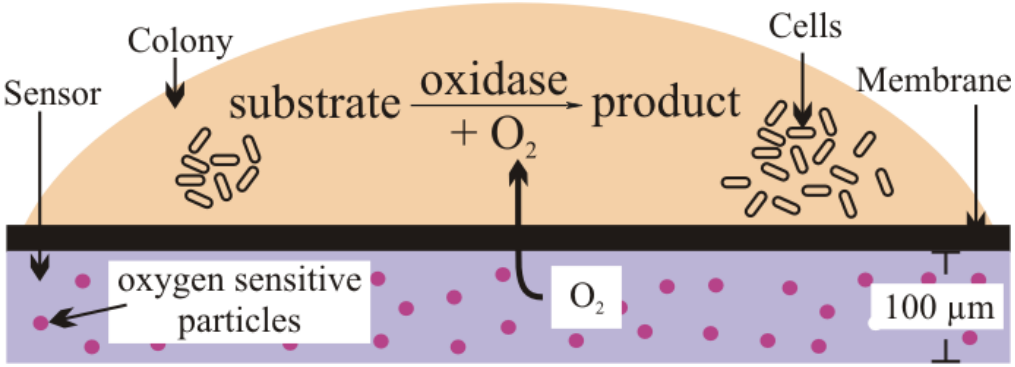




Features:

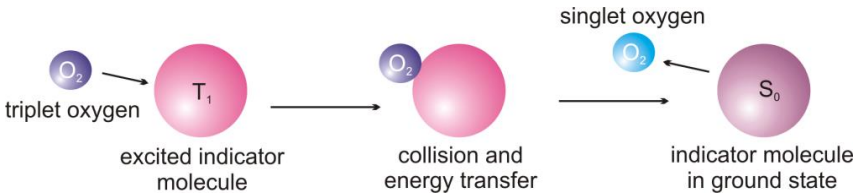
- Lower or higher buffer capacity allows adjustment to activity of the colonies.
- The pK of the indicator defines the sensitive pH window.
- Acryl derivatives of fluoresceine allow covalent linkage of the indicator to the matrix.

This system is usable for every enzymatic reaction which releases or consumes protons.

# Sensor design: $pO_2$-Sensor

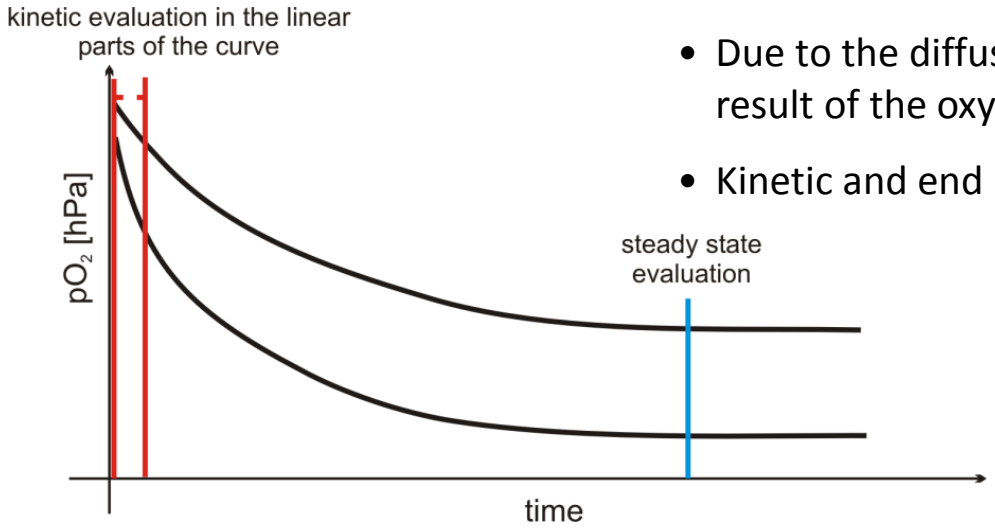



## Oxidase screening




Features:

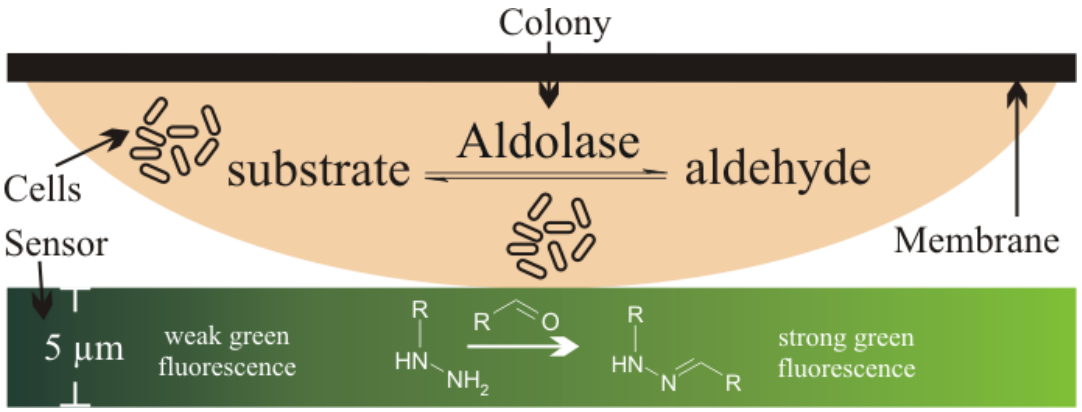
- Material of the particles allows adjustment to activity of the colonies.
- Due to the diffusion of oxygen in the sensor a steady state is the result of the oxygen consuming reaction.
- Kinetic and end point measurements are possible.

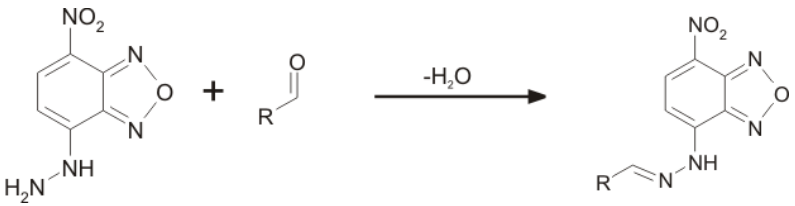


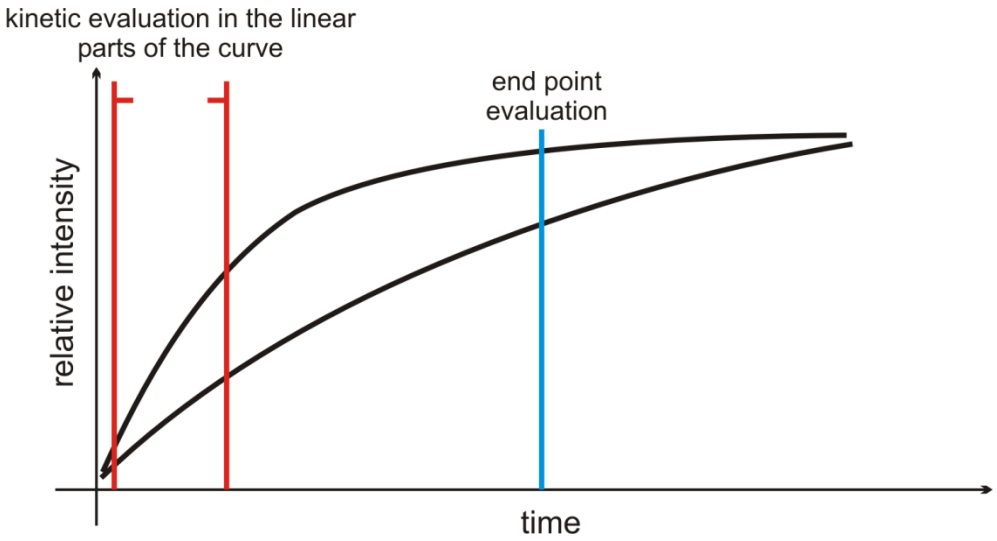

This system is usable for every enzymatic reaction which consumes oxygen.

# Sensor design: Aldehyde sensor

## Aldolase screening



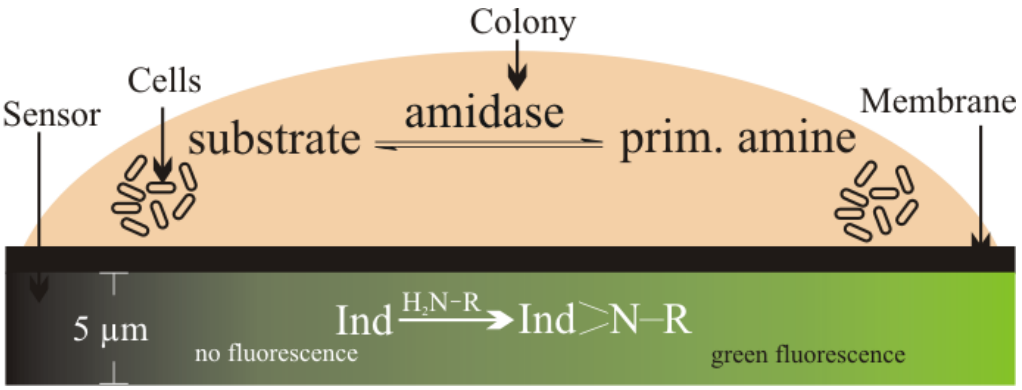

Features:

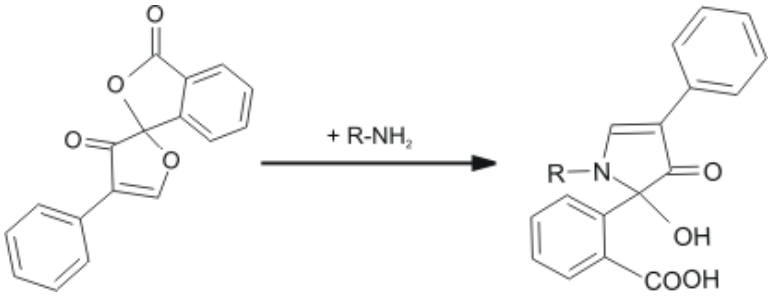
- Different matrices for different aldehydes.
- Screening in non-aqueous environment is possible.
- Incorporation in MTP should be possible.
- Other indicators have to be examined.

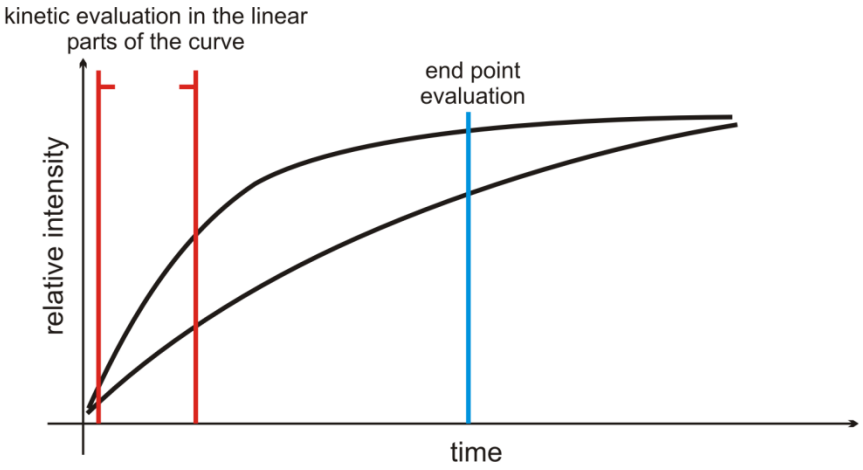
This system is usable for every enzymatic reaction which releases aldehydes.

# Sensor design: prim. Amine sensor

## Amidase screening




Features:

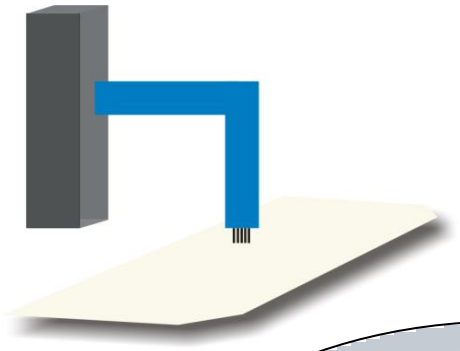
- matrix determines the screenable activity range.
- Sensor is easily adaptable for other amine indicators.
- Kinetic and end point measurements are possible.
- Incorporation in MTP is possible.
- Protease reactions are visualisable.

This system is usable for every enzymatic reaction which releases prim. amines.

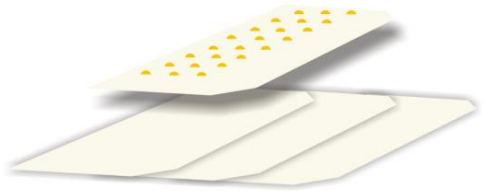
# Micro-colony Array Screening Platform

## Procedure
## for screening with micro colony chips

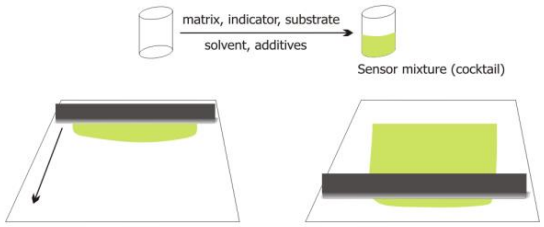
### Spotting



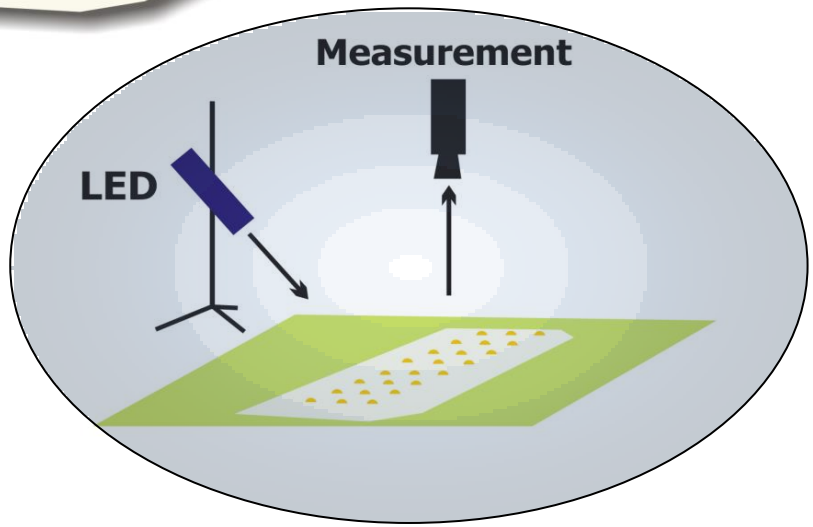
### Replication



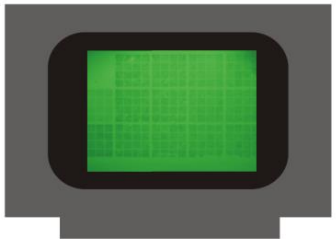
### Sensor concepts




### Measurement

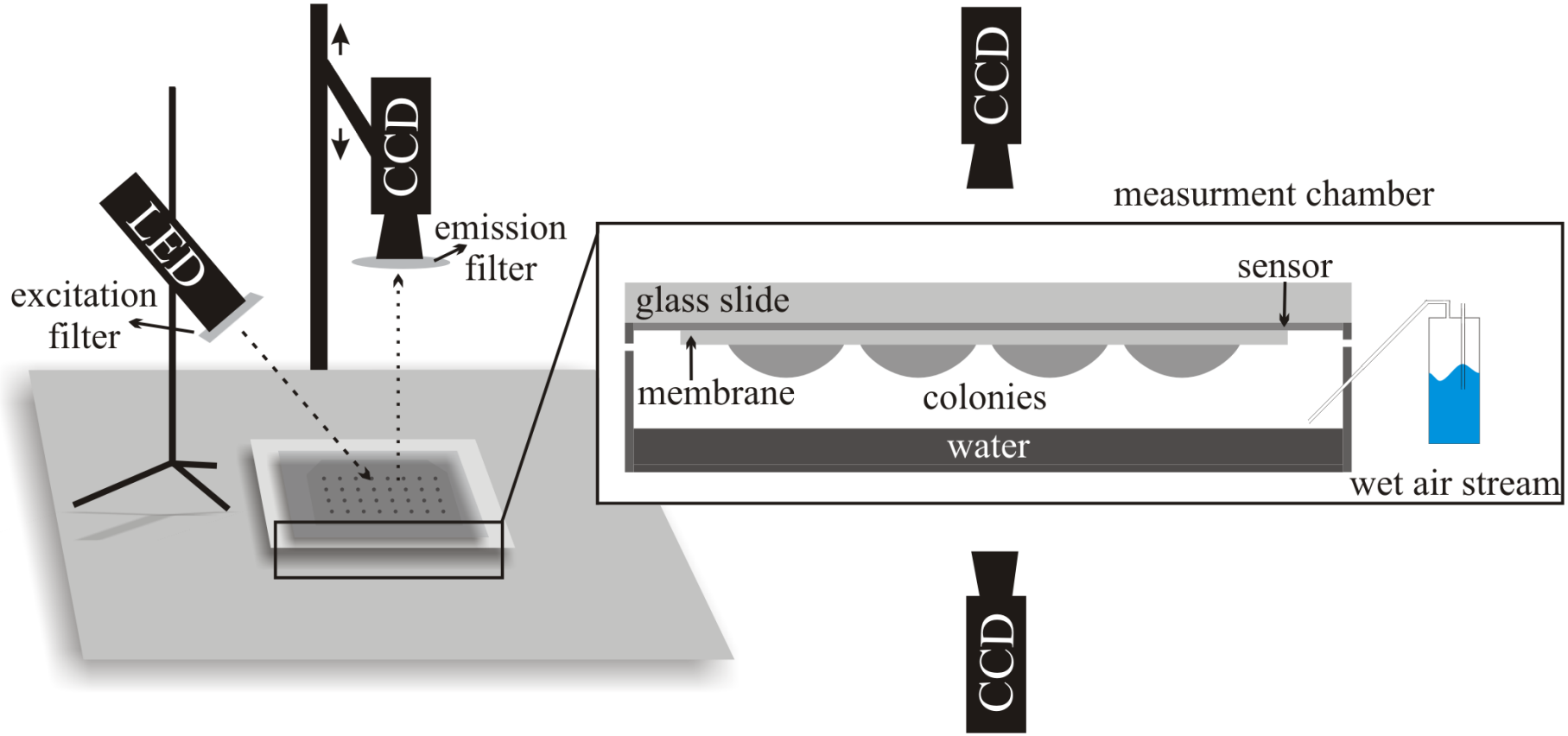



### Data acquisition



# Measurement – instrumentation

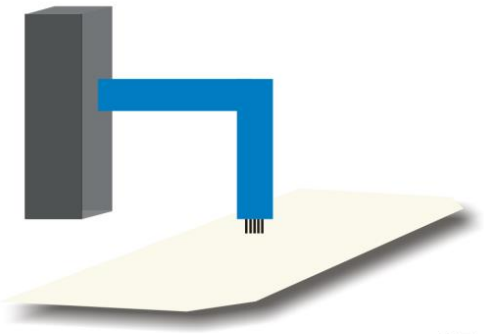
## Overview



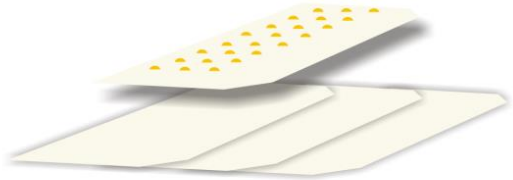
# Micro-colony Array Screening Platform

## Procedure
## for screening with micro colony chips

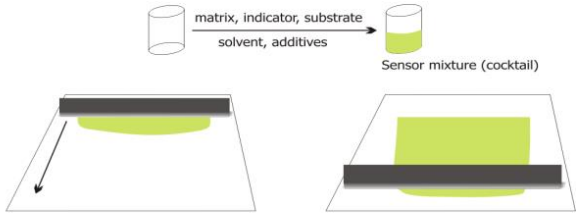
### Spotting



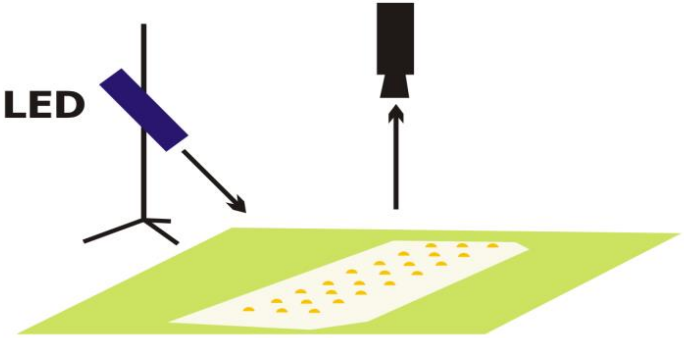
### Replication



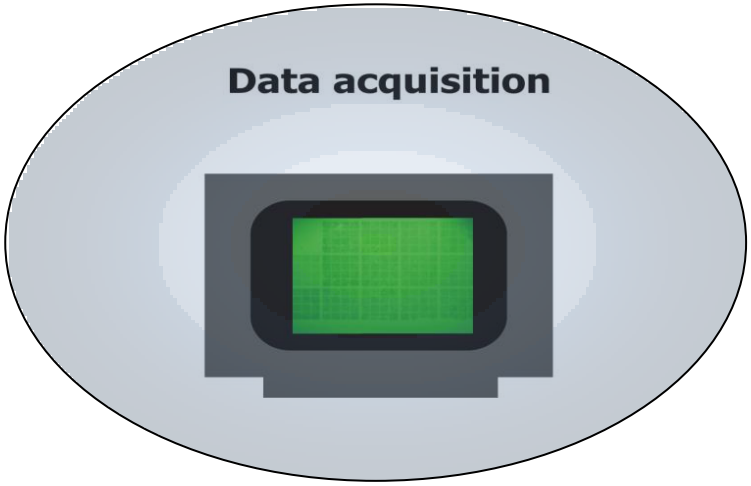
### Sensor concepts



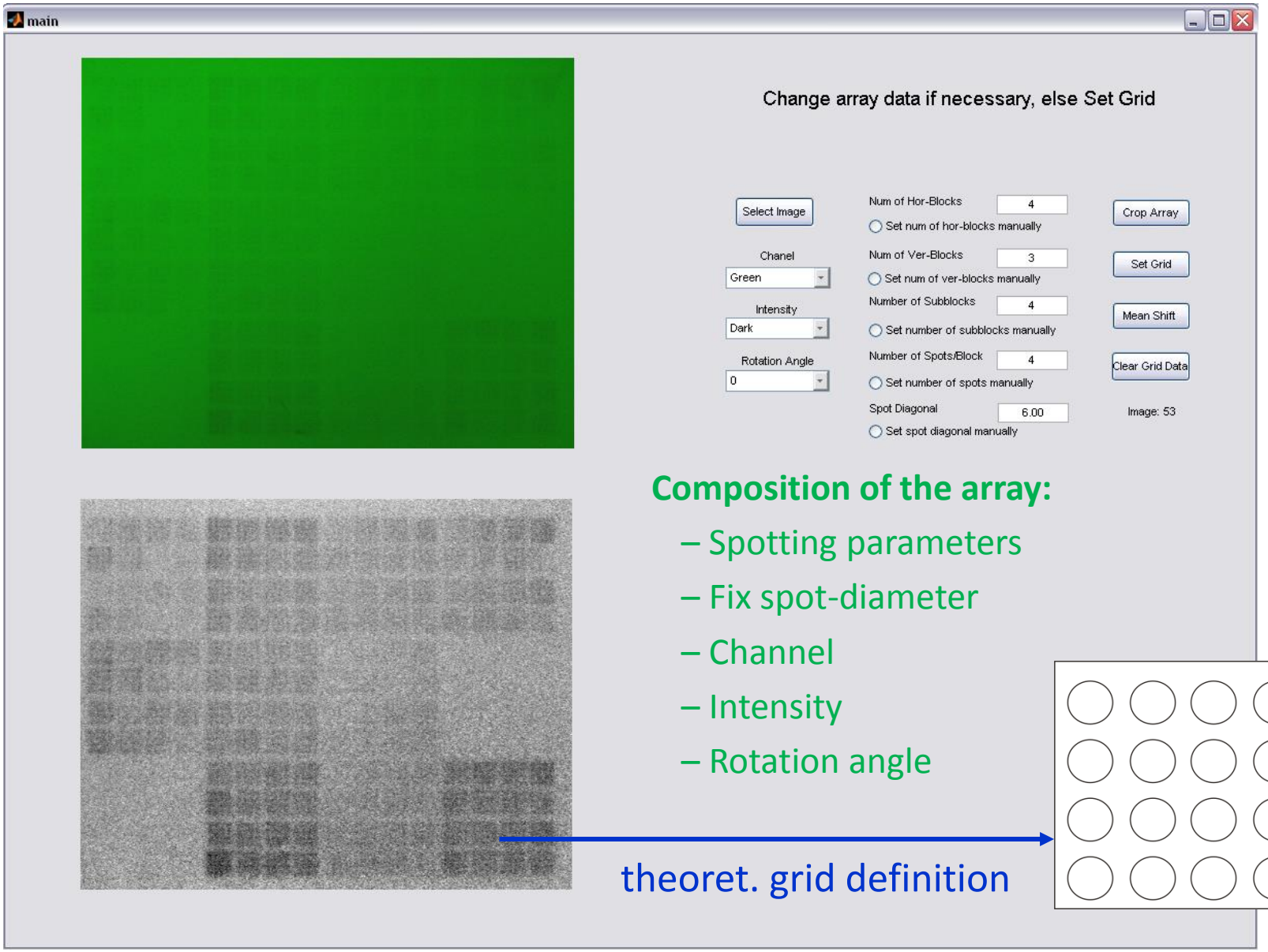

### Measurement




### Data acquisition



# Measurement – Data Evaluation manually




**Composition of the array:**

- Spotting parameters
- Fix spot-diameter
- Channel
- Intensity
- Rotation angle

theoret. grid definition

# Measurement – Data Evaluation automatically

… **normalise image for comparison**

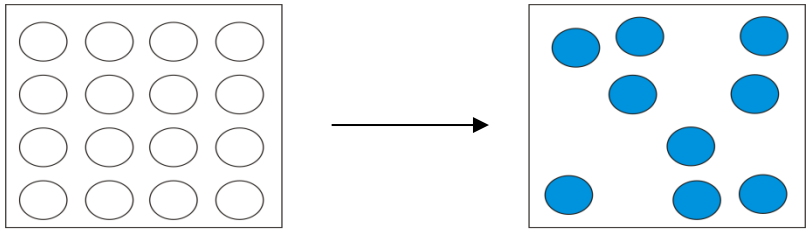
… **find the position of the spots**

- Corresponding block
- Situation inside the block

… **find the centre of the colonies**

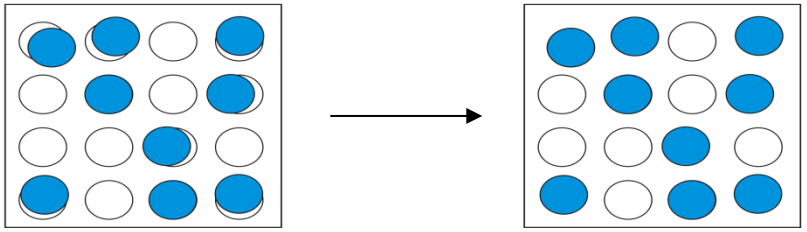
- Spot-content from centre and diameter

theoret. grid definition



real grid

Finding the centre of the colonies



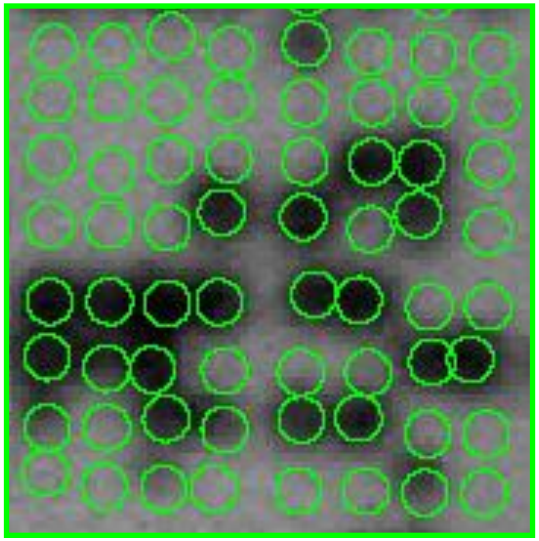
at found positions, calculation can be done

# Measurement – Data Evaluation

## Results

**Spot positions are …**

- clearly identified as long as there is a signal.
- approximated when there is no signal.
- missed if the real centre is too far from the first grid approximation.



**Colony chips with less than ~ 30% visible spots:**

- alignment of the theoret. grid with the real grid not possible.
- guide spots are not replicable.
- an image of the grid is needed – second camera for colony mage

# 7.5.15