

# I. Intro

## Anwendung MolBio:

- Krankheiten . K.O. Gene / O.E. Gene
- Züchtung von Pflanzen, Resistenzen
- Herstellung von Medikamenten, Enzymen, ...
- Vaterschaftstests
- phylogenetische Analysen

## gerichtete Evolution: Proteindesign

Beispiel: Invertase im Wein

Invertase: hydrolysiert Saccharose zu Glucose + Fruktose

→ im Wein:

- Zuckerzusatz steigert Alkoholgehalt
- Saccharose nicht vergärbar, Glucose schon

Problem: Glukose hemmt Produktion der Invertase  
(Selbstregulation)

Ziel: erschaffe Enzyme, die nicht durch Endprodukt inaktiviert werden

Methode: gerichtete Evolution

1. DNA Isolierung: Invertasegen

Input: Hefe mit Invertasegen

→ Zellmembran + Zellkern auflösen

→ Invertasegen suchen + isolieren

2. Invertasegen in E. coli geben

3. Mutagenese der E. coli

4. Enzymassay + Selektion

) wiederholen bis Ziel erreicht

## Phosphat - Problem

- Phosphat - Zyklus beinhaltet nur Land + Meer, NICHT die Atmosphäre
- exzessive Phosphat - Nutzung  $\rightarrow$  Vorräte schwinden

Lösungen:

- Phosphat - Recycling
- Pflanzen mit effizienterer Nutzung züchten

## II. DNA

### Struktur

DNA = Doppelhelix aus Nukleotiden  
 $\rightarrow$  verbunden über Phosphodiester-Brücken

- 2 komplementäre Stränge: antiparallel  
 $\hookleftarrow$  über Wasserstoffbrückenbindungen zusammen

### Nukleotide

= Pentosezucker  
+ Stickstoffhaltige Base  
+ Phosphatgruppe

} Nukleosid

### Pentosezucker

RNA: Ribose

DNA: Desoxyribose

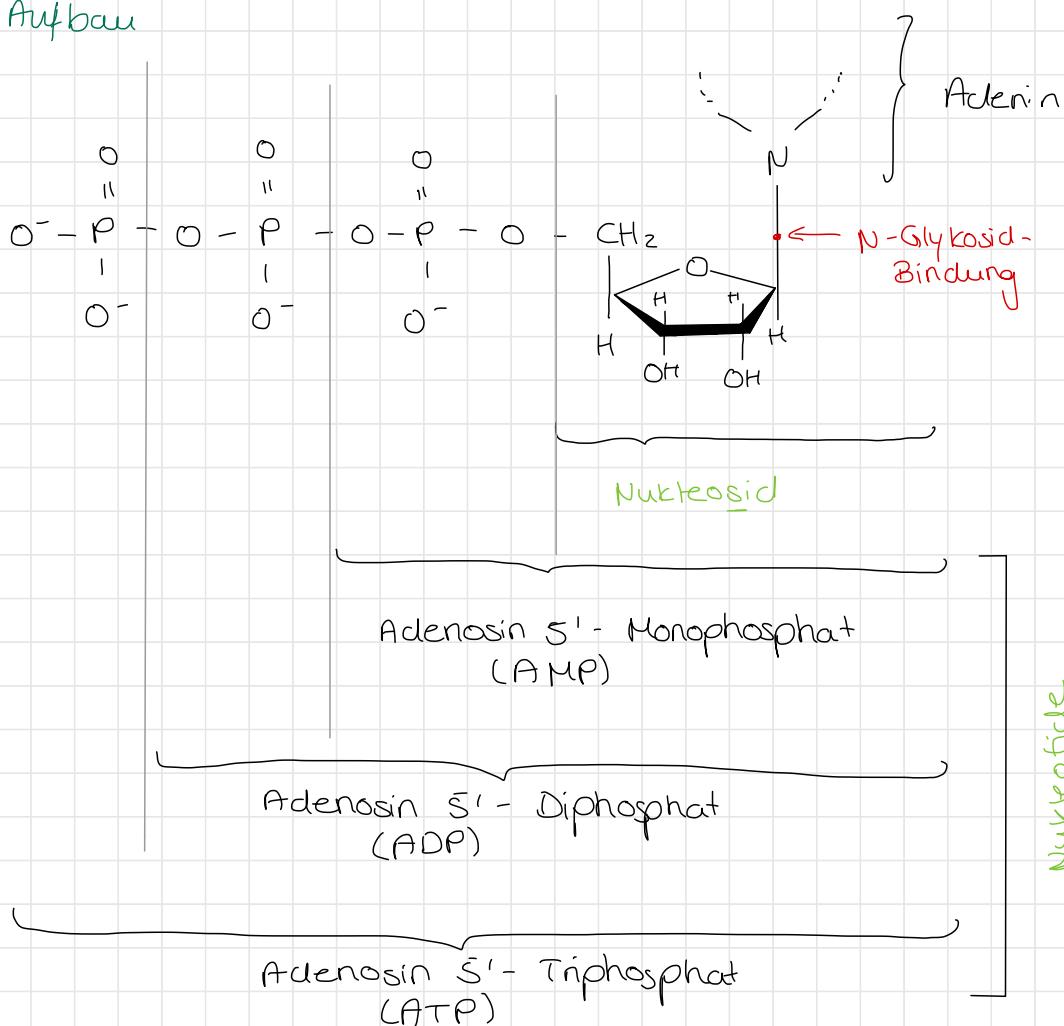
### Base

- Purine: Adenin, Guanin
- Pyrimidine: Thymin, Cytosin

RNA: Thymin durch Uracil ersetzt

$\rightarrow$  T hat am C-5-Atom Methylgruppe  
U OH

## Aufbau



## Basenpaarung

immer Purin + Pyrimidin (Breite stimmt nur so)  
 · immer C + G      } # Wasserstoff-  
                   A + T      } Brückenbindungen

## Chemische Eigenschaften

- polar, negativ geladen
  - 5'-Ende: freie Hydroxyl-/Phosphatgruppe an C-5'
  - 3'-Ende: freie Hydroxylgruppe an C-3'

- Salz stabilisiert Doppelhelix
- DNA Stränge: reversibel denaturierbar (trennbar)
  - zB Trennung durch Hitze / alkalische Lösungen
  - AT-Regionen zuerst getrennt (nur z. wass. br. bind.)
- Schmelztemperatur  $T_m$  = Temp. bei der 50% der Moleküle Einzelstrang, 50% Doppelstrang
  - proportional zum GC-Gehalt
  - prop. zur Ionenstärke / Salzkonzentration

## Gelelektrophorese

- Auftrennung geladener Teilchen nach Größe durch ein Gel mittels elektr. Feld

### Material:

- Plastikgefäß, mit Strom verbunden
- Puffer
- Salzlösung
- Gel (polymere Materialien)  $\xrightarrow{\text{Agarose für 200-50000 bp}}$   $\xrightarrow{\text{Polyacrylamid für 1-500 bp}}$
- Spannungsquelle
- Farbmittel

### Funktionsweise:

- DNA / RNA Moleküle = stark ionisiert bei neutralem pH, negativ geladen  
 $\rightarrow$  laufen zum  $(+)\text{-Pol}$
- $\rightarrow$  kleine = schneller

### Methodik:

1. Gel in Plastikbox mit Salzlösung geben
2. DNA Samples in Slots im Gel pipettieren  
 $\hookrightarrow$  Samples sind farbig markiert
3. Anlegen eines elektrischen Felds

## Markierung

= sichtbar machen im Gel

zB mit Ethidium-Bromid → interkaliert zwischen die Basenpaare der DNA

## Problem: Zirkuläre DNA

- in Prokaryonten
- in Mitochondrien, Plastiden (Endosymbionten-Theorie)

→ bilden Supercoils

⇒ supercoiled DNA läuft schneller im elektr. Feld

⇒ zuerst Ring auflösen vor Gelektrophorese

## Denaturierung:

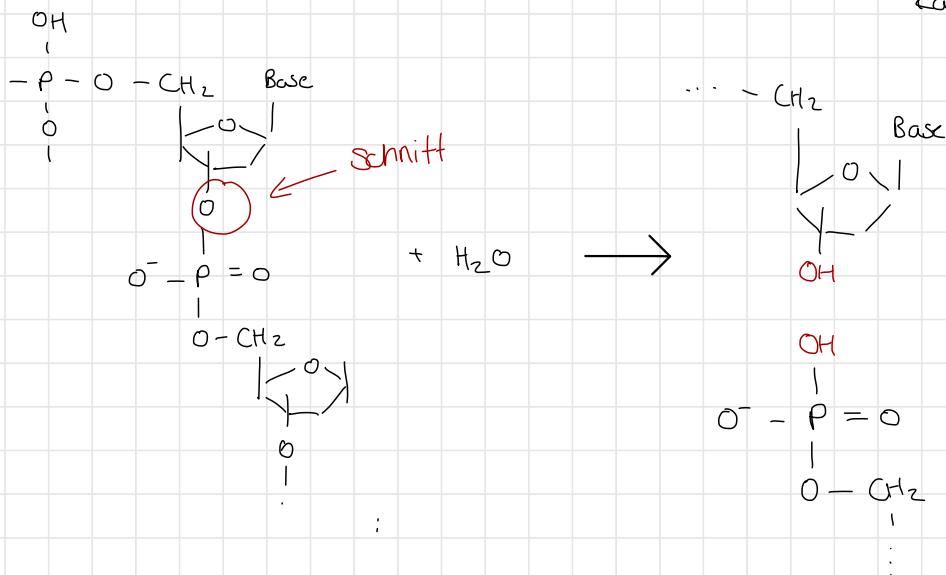
- zirkuläre DNA nicht denaturierbar durch Hitze/Alkali  
→ erst Einzelstrangbruch (Nick) einführen

### III. Restriktionsenzyme

- = Restriktionsendonukleasen
- Bakterielle Enzyme
- Erkennen + schneiden spezifischer 4-8 bp - Sequenzen der DNA

#### Restriktionsendonukleasen

= Enzyme, die die Hydrolyse des Phosphodiester-Rückgrats der DNA katalysieren



#### Eigenschutz

- Restr. enz. schneiden Fremd-DNA
- Schutz eigener DNA an spezifischer Sequenz: Methylasen

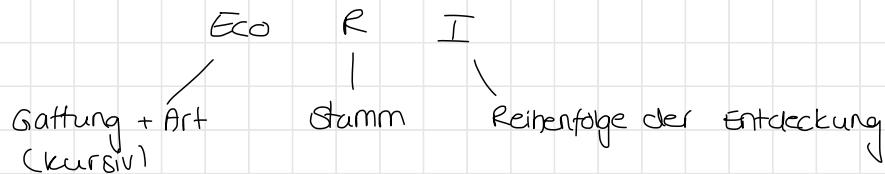
#### Methylasen

- = modifizierende Enzyme
- methylieren an best. Basen
- ⇒ Restr. enz. kann nicht mehr binden

⇒ je Restr. enz. gibt es Schutz-Methylase

= Restriktions-Modifikations-System

## Nomenklatur



## Restriktionskarte

- Restriktionsanalyse der DNA = Restriktionsfragment
- Muster (= Länge + Häufigkeit) spezifisch je Organismus  
    ≈ spezifische DNA
- sichtbarmachen durch Gelelektrophorese
- ⇒ falls Mutation in Basenpaar der Restr. enz. - Zielsequenz:  
    kein Schnitt / mehr Schnitte als normal  
    = verändertes Bild der Gelelektrophorese

## Lineare DNA

- n Schnitte = n+1 Fragmente

## Zirkuläre DNA

- n Schnitte = n Fragmente

} ⚠ Aufgabenstellung genau lesen

## Rare Cutters

- > selten schneidende Enzyme
- = haben lange Erkennungsseq.
- je länger diese ungewöhnlich ist zufälliger Schnitt

zB 4 Cutter: 1 Schnitt alle  $4^4 = 256$  Basen  
8 Cutter  $4^8 = 65536$  Basen

## Multi Cutters

- erkennen mehr als 1 Sequenz, zB HincII: 4 verschiedene

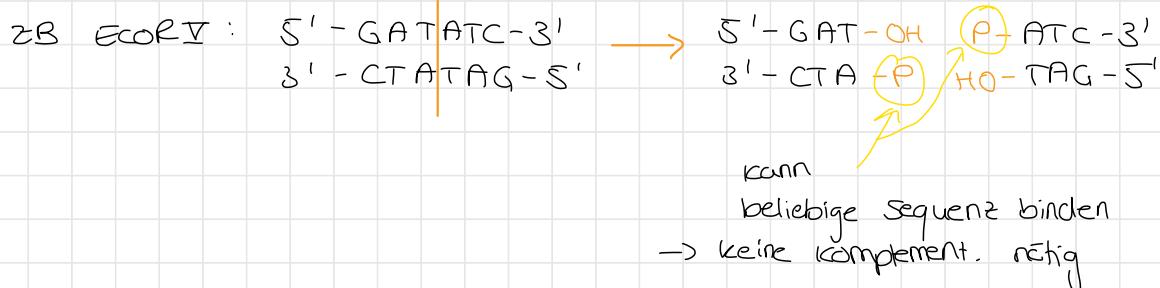
## Fragment - Enden

- blunt-ends oder sticky ends

! immer Enden dazu schreiben (3' / 5')

### blunt ends "stumpfe" Enden

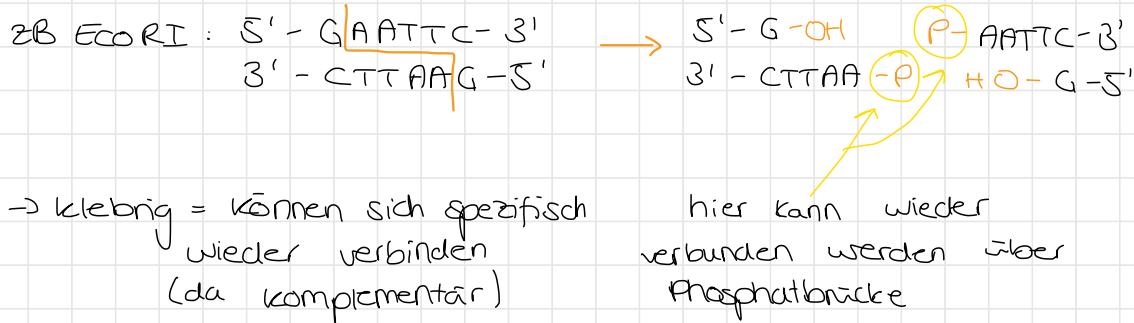
· entstehen durch vertikalen Schnitt des Restr.enz.



### sticky ends "klebrige" Enden

= cohesive ends

· entstehen durch Zick-Zack Schnitt des Restr.enz.



### blunt-ending

= abstumpfen der cohesive Ends

Ziel: stumpfes Ende das beliebige Ligierung erlaubt

2 Möglichkeiten:

- Polymerase = Füllen
- Exonuklease = Abschneiden

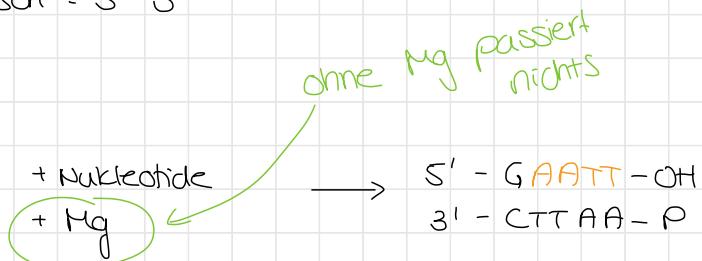
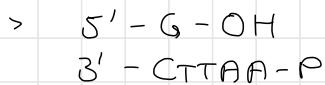
## Polymerase

- füllen Nukleotide in 5' - 3' - Richtung auf

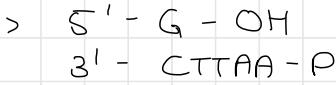
## Exonuklease

- DNA-Polymerasen können Exonukleaseaktivität haben
- beide Richtungen grunds. möglich
- ZB · Korrekturlesen = 3' - 5'
- Strang-Austausch = 5' - 3'

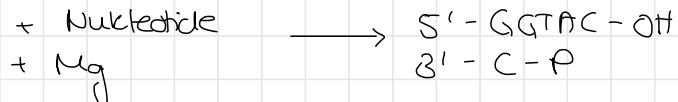
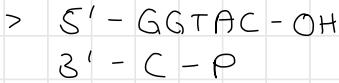
## Beispiele:



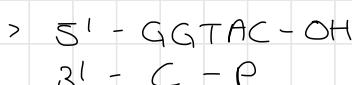
= **Polymerase**, da Nukleotide vor Abbau schützen  
Mg als Kofaktor vorhanden  
Richtung 5' - 3' stimmt



= **Exonuklease**, da keine Nukleotide  
Mg als Kofaktor vorhanden



= passiert nichts, da Nukleotide Exonuklease verhindern  
Richtung 3' - 5' für Pol geht nicht



= **Exonuklease**, da keine Nukleotide  
Mg als Kofaktor vorhanden

## Klenow-Fragment

- Exonuklease nur in 3'-5' - Richtung

## T4 DNA-Polymerase

- Exonuklease nur in 5'-3' - Richtung

## Restriktionsenzymklassen

### Typ I

- spezifische Erkennung, schneiden ssDNA weit entfernt von Erkennungsseq.
- Kofaktoren: Mg<sup>2+</sup>, ATP, S-Adenosylmethionine

### Typ II

- spezifische Erkennung, schneiden dsDNA nahe/innerhalb Erkennungsseq.
- Kofaktoren: iDR Mg<sup>2+</sup>  
→ typischerweise benutzt  
zB EcoRI, HindIII, NotI

### Typ III

- Schneiden dsDNA außerhalb Erkennungsseq.
- Kofaktoren: Mg<sup>2+</sup>, ATP

### Typ IV

- erkennen modifizierte (iDR methylierte) DNA

---

## Übung: Restriktionskarten

1. Linear oder Zirkulär?

2. Banden summieren + prüfen

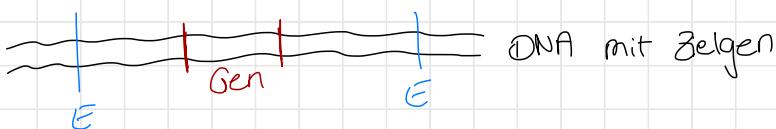
3. Restr. Karte Zeichnen

    ? 2 Enzyme schneiden NLE an der gleichen Stelle

4. am Ende nochmal ALE prüfen

## IV. PCR & DNA Klonierung

Bisher:



Restriktionsenzym schneidet Zielgen-Region aus DNA

Jetzt: Zielgen gezielt amplifizieren (=vervielfältigen)

### PCR

· in vitro DNA - Replikation

### Material

- template DNA
- 2 Primer (Forward- + Reverse-Primer)
- dNTPs (= Nukleotid)
- Puffer mit Mg<sup>2+</sup> (Kofaktor für DNA-Pol)
- hitzestabile DNA-Polymerase
- Thermoblock

### Funktionsweise

- 20-30 Zyklen
- nach ca 30 Zyklen: Plateau
- nach n Zyklen:  $2^n$  Moleküle

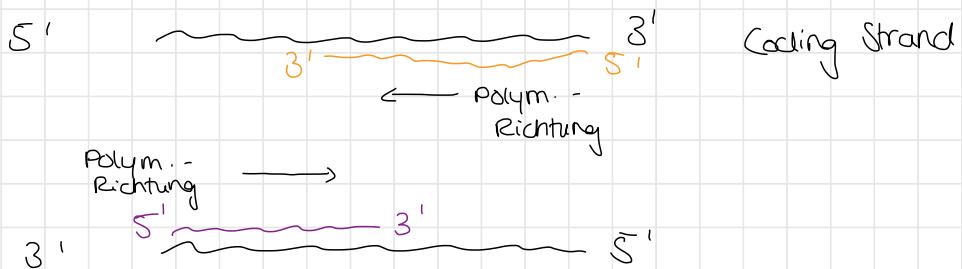
## Zyklus

1. Template Denaturierung  
= Trennung der Doppelstrände
  2. Primer annealing bei Temperatur  $T_m$   
-  $T_m$  abh. von den Primern (GC-Gehalt)
  3. Primer Extension = DNA Synthese  $T_s$   
- Temp. abh. von DNA-Polymerase  
- zB 72°C für Tag-Polymerase

→ nach n Zyklen nochmal 10 min bei  $T_s$  damit alle Strände vollst. synthetisiert

## Primer Annealing

- Reverse primer: bindet Coding Strand ( $5'$  -  $3'$ )
  - Forward primer: bindet  $3'$  -  $5'$  - Strand, Sequenz identisch zu Coding Strand



! Richtungen beachten  
o + dazu schreiben

## DNA Klonierung

- in vivo DNA-Replikation

→ amplifizierte geschnittene DNA Fragmente mittels Vektor vermehren

### Funktionsweise

1. DNA in Vektor ligieren
2. Ligierte Vektoren in Wirtszelle geben  
→ Replikation = Klonierung
3. Library Screening: Klone sortieren + suchen

### Vektoren

Voraussetzung:

- Origin of Replication
- Region für DNA insert (= Multiple Cloning Site)
- Resistenzgen für Selektion

### Beispiele

- Plasmide
- λ-Phagen
- Cosmide
- bacterial artificial chromosomes
- yeast artificial chromosomes

### Plasmide

- zirkuläre dsDNA, 1-200kb
- autonome Replikation (= unabh. von chrom. DNA)
- mind. 1 Kopie an Tochterzellen bei Zellteilung

## Insertion

= Einbringen Fremd-DNA in Vektor

## Funktionsweise

1. Vektor + Fremd-DNA mit Restr. Enzym schneiden  
=> lineare DNA-Fragmente mit OH<sup>-</sup> und PO<sub>4</sub><sup>3-</sup>

2. Ligation Vektor - Fremd-DNA mittels Ligasen

## Ligasen

### T<sub>4</sub>-DNA-Ligase

- Energieträger ATP
- verbindet blunt ends, sticky ends von RNA, DNA
- repariert Nicks
- Temp. 10 - 25°C (opt.: 16°C)

### E. coli-DNA-Ligase

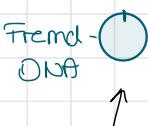
- Energieträger NAD<sup>+</sup>
- verbindet sticky ends von dsDNA
- Temp. 10 - 25°C (opt.: 16°C)

## Ligation

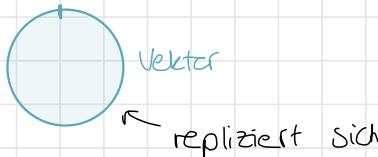
Fragmente nach Restriktionsverdau frei in Lösung  
=> versch. Ligationen möglich  
= versch. DNA-Stränge

## Selbstligation

- Vektor und Fremd-DNA ligieren sich selbst



↑  
sterbt, kann  
sich nicht  
replizieren



repliciert sich

## Lösung:

- Dephosphorylierung mit alkalischer Phosphatase (Enzym)  
= entfernt P von 5'-Enden
- ⇒ nur ein Fragment dephosph. (entw. Vektor oder Insert)

## Linker / Adapter

- kreieren Restriktionschnittstellen falls Enden nicht kompatibel
  - z.B. falls Vektor sticky end  
Insert blunt end

## Adapter

- baut sticky end an

## Linker

- baut kurze Doppelstrang-Sequenz an  
→ Restr.verdau mit gewünschtem Restr. enzym

## Transfer

- Übertragung Vektor → Wirtszellen

## Methoden:

- Hitzechock
- Elektroporation
- Mikroinjektion
- Gene Gun

## Hitzeschock

- mit  $\text{CaCl}_2$  kompetenten Zellen
- kurzer Hitzeschock öffnet Zellmembran an kleinen Löchern  
 → Unterdruck  
 → Vektor strömt nach innen  
 → Wirtsenzyme binden ORI und bewirken Replikation

## Elektroporation

- mit elektrokompetenten Zellen
- Kurzer Elektroschock sorgt für kurze Porenbildung in Zellmembran  
 → kurzzeitig permeabel
- ⊕ mehr Kolonien als Resultat als bei Hitzeschock  
 = höhere Effizienz

## Selektion

= prüfen ob Vektor in Wirtszelle aufgenommen

## Plasmidklonierung

- erlaubt Isolierung von DNA-Fragmenten aus komplexen Mischungen

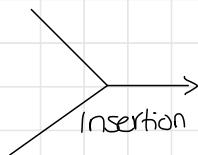
Plasmid

Vektoren



DNA

Fragmente



→ in Wirtszelle  
+ auf Agarplatte

⇒ Kolonien auf Agarplatte  
eine Kolonie = ein Klon

## blue - white screening

- positive Selektion

### Ziele:

1. Teste ob Transfer Vektor  $\rightarrow$  Wirt erfolgreich  
 $\rightarrow$  Selektionsmarker
2. Teste ob Insert DNA-Fragment  $\rightarrow$  Vektor erfolgreich  
 $\rightarrow$  lacZ - Gen

1) Selektionsmarker, zB AmpR

$\rightarrow$  gebe Wirte auf Agarplatte mit Ampicillin

$\rightarrow$  nur Wirte mit Vektor überleben

2) lacZ +  $\beta$ -Galactosidase

- falls Insert erfolgreich: Gen kaputt

$\Rightarrow$  weiße Färbung

- nicht erfolgreich: Gen intakt

$\Rightarrow$  blaue Färbung

- Lac - Operon ( $E. coli$ ) metabolisiert Laktose in Abwesenheit von Glukose

$\rightarrow$  verwende Agarplatte mit X-Gal statt Laktose

$\Rightarrow$  Lac - Operon spaltet zu Galactose + Stoff der blau wird

## Km - Resistenz

= negative Selektion

• Vektor-Wirt - Selektion mit Ampicillin wie oben

• Insert DNA - Fragment in Km - Resistenzgen

→ 1. gebe Wirtszellen auf Medium ohne Km

2. übertrage Kolonien mit Stempel auf Medium mit Km  
⇒ Zellen mit erfolgreichem Insert sterben  
(da Insert Km - Resistenz kaputt macht)

3. Selektiere die toten Kolonien von Medium 2 in Medium 1

→ das sind die mit erfolgreichem Insert

## Genbanken

• # benötigter Klonen für genomische Bibliothek:

$$N = \frac{\ln(1-P)}{\ln(1-\frac{a}{b})}$$

• P = Wkt., dass jedes Gen vorhanden (zB 95%)

• a = Ø Größe der DNA-Fragmente

• b = Ges.größe Genom

# Klonen

Spezies	Genomgröße	17kb Fragmente	35kb - Frag m.
E. coli	4 639 kb	820	400
Bäckerhefe	12 520 kb	2 200	1 075
Mensch	3 000 000 kb	535 000	258 000
Mais	5 000 000 kb	900 000	450 000

## Plasmid - Genbank

- E.coli Transformation wenig effizient
  - wenige transformierte Kolonien pro Platte ( $10^6$  Plaques / µg DNA)
- => Klonierung großer Genome unpraktisch

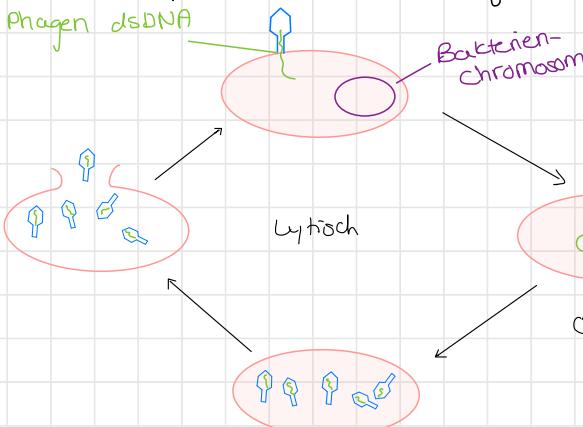
## $\lambda$ -Phagen - Genbank

### $\lambda$ -Phage

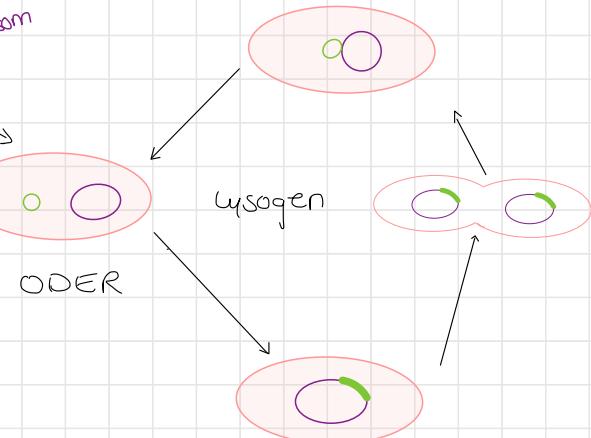
- Kopf mit viraler DNA
- Schwanz zur E.coli - Infektion

### Zyklen

Lytisch = Vermehrung im Wirt,  
töten des Wirts  
=> freisetzen vieler Phagen



Lysogen = Insertion in wirt-  
chrom. DNA, passive  
Replikation



### COS-sites

- Replikation  $\lambda$ -DNA im Wirt erzeugt concatemer = langes multimeres DNA-Molekül
- Multimere getrennt durch cos-sites

- Cos-sites = flankieren λ-DNA
  - NusA und A Proteine erkennen Cos-sites
  - + verpacken DNA dazwischen

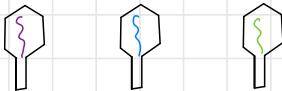
- (+) Insert bis zu 20kb Fragmente möglich
- effizient: je Phage  $> 10^6$  neue Phagen  
 $\Rightarrow 10^9$  Plaques / µg DNA
- (-) noch zu ineffizient für große Genome

## Cosmide - Genbank

- Cosmid-Klonierung = Phage - Plasmid - Vektoren
- Konkat. DNA-Fragmente mit Cosmid-Vektor:



→ durch Cos-sites: versch. Fragmente = versch. Phagen



→ Infektion von E. coli  
 → ...

- (+) Insert bis 45kb möglich  
 $\Rightarrow$  einfache Klonierung größerer Genome  
= effizient

## BAC - Genbank

BAC = bacterial artificial chromosome  
= Plasmid - Vektor mit F-Faktor

· F-Faktor ermöglicht Gentransfer

(+) · Insert bis 300kb möglich

## YAC - Genbank

YAC = yeast artificial chromosome  
= art. chrom. mit Telomer, Centromer, Repl. Ursprung  
für Hefezellen

(+) · Insert bis 300kb möglich

(-) · Instabiler als BAC durch häufige Rekombination

## V. Library Screening

Ausgangspunkt: Gen-Bibliothek = Pool aller gDNA Klonen

genomische DNA  
↓

Ziel: finde spezifisches Gen (= Klon in Gen-Bibliothek)

Lösung: Sonden - Käder, der bestimmten Klon findet

### Sonden

#### Antikörper-Sonden

- für Vektor mit Insert das exprimiert wird, weil es hinter Promoter kloniert
- => mRNA translatiert in Protein
- => Antikörper bindet Protein

### Mechanismus

- codierte Proteine der Gen-Bank müssen exprimiert werden

#### Phagen-Bibliothek:

DNA → mRNA → Protein

{ ↑ |

· kopiert in cDNA |

· Ligation in Phagen Vektor |

= Genbank mit cDNA |

---' muss jetzt  
da hin  
kommen

=> brauchte Promoter

1. Schneide  $\beta$ -Galactosidase - Gen mit EcoRI
2. bringe cDNA mit linker an sticky ends
3. Ligase
4. Durch lac-Promotor: Fusions-Protein ( $\beta$ -Galact. + cDNA) wird exprimiert

## Detection

1. Proteine auf Platte mit Bakterien-Rasen geben
2. Nitrozellulosefilter drauf
3. entfernen  
→ Proteine binden an Filter
4. Filter mit Primär-Antikörper inkubieren
5. Waschen (wäscht unspez. / schwache Bindungen aus)
6. Filter mit Sekundär-Antikörper inkubieren  
(binden an Primär-Antikörper)  
→ Sek.-Antikörper ist markiert
7. X-Ray: sichtbar machen der gesuchten Plaques auf Filter

## DNA / RNA - Sonden

- markiert mit  $^{32}\text{P}$  / DIG zum sichtbar machen
- für Vektor mit DNA/cDNA insert ohne Promoter
  - => keine Expression, nur Replikation
  - => Sonde hybridisiert mit gesuchtem Klon

## Detection

durch Membran-Hybridisierung

## Southern Blot

1. Gelektrophorese = Trennung der Bibliothek nach Größe
2. Gel in starke Launge: denaturiert dsDNA → ssDNA
3. ssDNA auf Nylonmembran übertragen = Blotting  
=> Nylonmembran enthält alle DNA-Stücke aus der Probe
4. ssDNA auf Membran fixieren
5. Membran mit markierter Sonde über mehrere Stunden hybridisieren

## 6. Sichtbar machen

- => falls Sonde mit Fragment hybridisiert: sichtbar im X-Ray  
= Zielfragment war dabei

## Sonden: Oligonukleotide

- mind. 20 bp lang  $\rightarrow$  je länger desto besser  
= geringere Wkt dass zufällige Bindung

### Quellen:

- bekannte, sequenzierte DNA-Fragmente  
zB ESTs = partiell komplementär zu Zielgen
- heterologe Sonde  
= verwandte Gene aus anderen Spezies
  - ⊖ nicht zwingend ausreichend homolog für Bindung
  - ⊕ je nach Komplementarität: Temp. der Hybridisierung variieren + Hybrid-Lösung sollte leicht denaturierend sein um Selbsthybr. der Sonden ssDNA zu verhindern
- von Proteinsequenz
  - 7 spezifische AAs nötig (= 21 bp)

## Proteinsequenzen

- Sequenzierung durch Edmann-Abbau oder MALDI-TOF

## Edmann-Abbau

- je Runde: 1 AA am Aminoterminal des Proteins abspalten  
+ bestimmen

- (+) · Reihenfolge der AAs bekannt

- (-) ·  $\leq 60$  AA bestimmbar

## MALDI-TOF

### · Massenspektrometrie

1. Verdau Protein mit Protease
2. Vermischen mit Matrix
3. UV-Bestrahlung  $\Rightarrow$  Ionisierung
4. Beschleunigung pos. Ionen im elektr. Feld  
 $\rightarrow$  Flugbahn + time-of-flight abh. von AA

- (+) · beliebig große Proteine sequenziierbar
- (-) · Reihenfolge der AA unbekannt

Problem: Degeneration des genetischen Codes

· AA durch versch. Codons codiert

$\Rightarrow$  Proteinsequenz nicht eindeutig in RNA übersetzbbar

Lösung: wähle AA-Sequenz mit möglichst geringer Degeneration

zB Leu, Arg, Ser (6 Codons) vermeiden!

$\Rightarrow$  wähle aus AA-Sequenz 20 Basen

(20 heißt ich kann 1 Base weg lassen)

Übung: Oligonukleotid-Sonde

Frage: 2 AA-Sequenzen gegeben, welche ist besser geeignet als Sonde?

P1: Asp - Phe - Gly - Lys - Asp - Tyr - Tyr - Ala

P2: Phe - Tyr - Ile - Asp - Lys - Phe - Gln - Val

## Lösung:

1) Degeneration aufschreiben + ausrechnen:

P1: Asp - Phe - Gly - Lys - Asp - Tyr - Tyr - Ala  
 GAU UUU CGA AAA GAU UAU UAU GC U  
 C C C G C C C A  
 U C  
 G G

$$= 2^1 \cdot 2^1 \cdot 2^2 \cdot 2^1 \cdot 2^1 \cdot 2^1 \cdot 2^1 \cdot 2^1$$

Y 8 AAS

→ brauche nur 7

=> streiche entweder vorne oder hinten

=> streiche Alia (= mehr Neg. als Asp vorne)

$$= 2^1 \cdot 2^1 \cdot 2^2 \cdot 2^1 \cdot 2^1 \cdot 2^1 \cdot 2^1 = 2^8$$

→ brauche nur 20 Nukleotide

=> Striche entweder erstes oder letztes Nukleotid

=> streiche letztes

$$= 2^1 \cdot 2^1 \cdot 2^2 \cdot 2^1 \cdot 2^1 \cdot 2^1 \cdot 1 = 2^7$$

→ Für P1 codieren  $2^7$  mögliche Sequenzen

p2: analog zu p1:

P2: Phe - Tyr - Ile - Asp - Lys - Phe - Gln - Val

UUU	UAU	AUU	GAU	AAA	UUU	CAA	GUU
C	C	C	C	G	C	G	A
		A				C	G

$$= 2^1 \cdot 2^1 \cdot 3 \cdot 2^1 \cdot 2^1 \cdot 2^1 \cdot 2^1 \cdot \cancel{2^1}$$

→ Streiche auf 7 AS

$$= 2 \cdot 2 \cdot 3 \cdot 2 \cdot 2 \cdot 2 \cdot \cancel{2}$$

→ Streiche auf 20 Nukleotide

$$= 2 \cdot 2 \cdot 3 \cdot 2 \cdot 2 \cdot 2 = 3 \cdot 2^5$$

$\Rightarrow 3 \cdot 2^5 < 2^7 \Rightarrow$  Peptid 2 besser geeignet, da Degeneration geringer

Frage: Entwerfe Primer (Forward, Reverse) für Peptid

P: Ile - Leu - Ile - Asn - Cys - Gln

Lösung: 1 mögliche Codon-Sequenz wählen

→ ATT CTT ATT AAT TGT CAA

- Forward-Primer (= Codon-Sequenz):

5' ATT CTT ATT AAT TGT CAA 3'

- Reverse-Primer (= komplementär ↑)

3' TAA GAA TAA TTA ACA GTT 5'

→ in 5'-3'-Richtung angeben!

= 5' TTG ACA ATT AAT AAG AAT 3'

## Sonden - Markierung

### Oligonucleotid - Markierung

- markiere ssDNA mit  $\gamma$ - $^{32}P$  unter ATP Verbrauch mittels T4-PolyNucleotid-Kinase
- tausche 1 P im Backbone durch  $^{32}P$  aus

### End - Labeling

- markiere dsDNA durch anhängen eines markierten Adenin

#### 1. Dephosphorylierung

= ersetzt P an 5'-Enden durch OH

#### 2. T4-PolyNucleotid-Kinase: spalte $\gamma$ - $^{32}P$ von $\gamma$ - $^{32}P$ -ATP ab → an 5'-Enden anbauen

## Nick Translation

#### 1. Führe Strangbrüche in DNA ein (DNase I) = Nicks

#### 2. DNA-Pol I / Klenow + DNA-Ligase reparieren Nicks → stelle Nukleotide zur Verfügung - normales T, C, G - markiertes A: hat $^{32}P$

## In-vitro Transkription

#### 1. Trenne dsDNA - Stränge

#### 2. Lasse Klenow transkribieren

→ stelle dNTPs zur Verfügung mit  $\alpha^{32}P$ -dATP / DIG

## PCR - Labeling

1. Trenne dsDNA Stränge
2. Lasse Tag-Pol transkribieren  
→ stelle dNTPs zur Verfügung mit  $\times^{32}\text{P}$ -dATP / DIG

## Nichtradioaktive Markierung

- mittes DIG (aus Fingerhut) / Biotin
1. DNA -Labeling mit Nukleotiden, die Reporter- Gruppe tragen
  2. Koppelung Affinitäts- Gruppe + Marker -Gruppe
  3. Vermischen  $\Rightarrow$  Affinitäts- Gruppe bindet Reporter
  4. Sichtbar machen durch Licht

## DIG:

- Reporter- Gruppe: DesOxigenin (DIG)
- Affin. - Gruppe: DIG - spezifischer Antikörper gekoppelt mit alkalischer Phosphatase
- Marker: CDP-star  $\rightarrow$  löst Farbstoffumschlag aus bei Koppelung

# VI.

## Bioinformatik

### Datenbanken

- Sequenzdatenbanken
  - = Nukleotide, Proteine, ...
- Strukturdatenbanken
  - 3D - Proteinstruktur
- Pathway - Datenbanken
  - Stoffwechselwirkung, regulatorische Pfade
- Wissensdatenbanken
  - strukturelle + funktionale Info aus vielen Quellen

### Anforderungen

- Erreichbarkeit
- Widerspruchsfrei
- Aktualität
- gute Dokumentation
- Qualität der Daten
- Unterstützung gängiger Formate
- Flexibilität im Design

### Probleme

- Redundanzen (Kopien, Varianten)
- Qualität
- automatische Annotation nicht immer brauchbar

### Sequenzanalyse

#### Alignment

- local = finde Teilstück
- global = aligne gleichlange Stücke
- multiple = mehrere Stücke

#### Kosten

- Indels, Matches, Mismatches, Gaps Kosten zuweisen

## BLAST

Basic Local Alignment Search Tool

- keine optimalen Ergebnisse (heuristisch)

Schritte:

- 1) Seeding
- 2) Extension
- 3) Evaluation

### 1) Seeding

- Trenne Query in kleine Stücke
- Aligne Stücke an Reference
- Lücken

Query: ADE RQLSYT → ADE RQL SYT

Reference: ADEGLSYT

→ AOE GLSYT → übrig: GL / RQL  
 AOE SYT

→ AOE - GLSYT  
 AOE RQLSYT

### 2) Extension

- ausgehend von Seed: erweiterte Alignment in beide Richtungen
- => High-scoring-Sequence-pairs

### 3) E-Value

$$E \approx 1 - e^{-P(S > X) d} \quad \leftarrow \text{DB Größe}$$

- = Erwartete Anzahl, wie oft unverwandte DB-Sequenz zufällig score S > X hat

## NCBI - Datenbank

- BlastN : Nukleotid - Nukleotid Vergleich
- BlastX : translatierte Nukleotidsequenz - Protein Vergleich
- BlastP : Protein - Protein Vergleich

→ neue / unbekannte Sequenzen : anfangen mit BlastX / BlastP  
⇒ Protein kann von mehreren Seq. codiert werden  
= mehr matches als eine spez. Sequenz

## Phylogenetische Bäume

- branches = evolutionäre Abstammungen
- branch lengths = evolut. Zeit zw. 2 Knoten  
in substitutions per sequence site

## Aussagen

- Funktionen unbekannter Gene
- Zuordnung Gen → Organismus
- Ablauf Evolution
- Neofunktionalisierung von Duplikationen

## Genomik

### Vergleichende Genomik

- = Vergleich ≥ 2 Genomsequenzen  
→ finde gemeinsame Eigenschaften

Ziele:

- potentielle Funktionen für unbekanntes Gen bestimmen
- finde konservierte Genabschnitte
  - ≈ gleich bei allen Nachfahren  
= wichtige Funktionen

## Homologiesuche

- Homologie = ähnliche Sequenzen in versch. Organismen / Geweben die für Protein mit gleicher Funktion codieren  
→ gleich durch gemeinsame Abstammung

orthologe

= homolog in versch.  
Organismen

paraloge

= homolog im selben  
Organismus mit  
versch. Funktionen

- Analogie = ähnliche Funktion, aber keine gemeinsame Abstammung (konvergente Evolution)

## Syntenie

- = Gene in gleicher relativer Position auf 2 Chromosomen
  - zB versch. organismen
- vergleiche Genkarten versch. Organismen / Spezies
  - nahe verwandt: gleiche Anordnung

### VII. Sequenzierung

#### Kettenspaltung - Methode

- DNA Restriktions fragmente : Enden (radioaktiv) markieren  
→ Spaltung nach spez. Molekülen

zB: 4 tubes:

- T1:	zerstört	G	?	Zerstört nicht alle → probabilistisch!
- T2:		A, G		
- T3:		T, C		
- T4:		C		

→ Gelektrophorese aller Tubes = 4 Bahnen

+ ablesen:

- falls G: T1 und T2 zeigen Bande
- A: T1 keine, T2 hat Bande

:

(-) Zu kompliziert + aufwändig

#### Kettenabbruch - Methode

= Synthesemethode = neue DNA - Kette wird gebaut

→ verwendet dNTPs

↑ eine  $\beta'$ -Hydroxylgruppe fehlt

→ kann keine Phosphodiesterbindung mit nächstem dNTP bilden

#### Methode

• DNA - Fragmente + 5' markierter Primer + DNA - Polymerase + dNTPs  
→ 4 Tubes:

• T1: + ddATP	?	Stoppen Synthese sobald DNA - Fol. dNTP einbaut
• T2: + ddGTP		
• T3: + ddTTP		
• T4: + ddCTP		

- Denaturierung
- Gelelektrophorese
  - basengenaue Trennung!
  - dünne Gels, Hochspannung, ...
- ablesen: von unten (kurz) nach oben (lang)

### Markierung

- Alternative zum Ablesen:
- markiere dNTPs mit 4 versch. Farben
- detektiert automatisch im Gel mit Laser + Detektor

## Next Generation Sequencing

### Pyrosequencing

Grundlage: Nachweis des Pyrophosphats der während Sequenzierungsreaktion produziert

- wenn DNA-Pol. dNTP einlöst wird  $\text{PP}_i = \text{Pyroph. ftei}$
- ⇒ Menge  $\text{PP}_i$  prop. zur # angehängter dNTPs

Reaktion:

1.  $\text{PP}_i + \text{APS} \xrightarrow[\text{Sulphydrylase}]{\text{ATP-}} \text{ATP}$
2.  $\text{ATP} + \text{Luciferin} + \text{O}_2 \xrightarrow{\text{Luciferase}} \text{Oxyluciferin} + \text{AMP} + \text{PP}_i + \text{CO}_2 + \text{Licht}$ 
  - Detektion mit CCD - Kamera
3. Apyrase: Abbau unverbrauchter dNTPs, ATP

### Methode

- dNTPs werden sequentiell zugesetzt
- nach jeder Zugabe: Abbau mit Apyrase (= Reinigung)

→ Peak nur wenn dNTP an dieser Stelle eingebaut

- falls K gleiche Nukleotide hintereinander in DNA:  
K - mal so hohe Lichtemission (~ Peak im Diagramm)

## 454 Sequencing

Idee: gleichzeitige Pyroseq. vieler Fragmente

1. DNA Extraktion + Fragmentierung
2. Nebulization = DNA-Scherung
3. Adapter Ligation: A - Adapter an einem Ende → Bindung  
B anderem → Primer für DNA-Pol.
4. Bindung an Latex-Kugelchen mittels Adapter  
→ 1 Fragment je Kugel
5. Vermischen mit Wasser-Öl Emulsion  
→ jeder Tropfen enthält 1 Kugel
6. PCR-Amplifizierung auf jeder Kugel  
→ Kugel hat jetzt viele identische Fragmente
7. Auf Platte geben: Platte hat kleine Löcher, je Loch passt nur 1 Kugel
8. Enzyme für Reaktion zugeben
9. nacheinander: dNTP Zugabe → Reaktion abwickeln  
→ saubern
10. Sequenz zusammensetzen durch Finden überlappender Bereiche

(+) · 1 Durchlauf > 100 Mio Basen in < 8h

## 3rd Generation Sequencing Nanopore Sequencing

- Proteinpore in nicht-leitfähiger Polymermembran
- Anlegen einer Spannung  
→ Ionenfluss durch Pore

→ Helikase führt DNA/RNA-Molekül durch Pore  
⇒ je nach Base: Ionenfluss ändert sich  
⇒ messe Ionenstrom während Translokation

## Genomsequenzierung Genom

= gesamtes gen. Material einer Zelle

## Genomik

- strukturell = Entdeckung neuer Gene + Lokalisation  
→ Genkarten
- funktionell = biolog. Funktionen der Gene + Regulation, Produkte
- vergleichend = Vergleich Gen- / Proteinseq. versch. Genome  
→ evolutionäre, funktionelle Beziehungen

## Genomkartierung

2 Arten von Genkarten:

- genetische Genkarte (in cM) über Rekombination  
· je größer Abstand desto höher Wkt für Crossing-Over
- physikalische Genkarten (in bp) über Sequenzierung

## Modellorganismen

- Drosophila für Invertebraten
- Mäuse für Säugetiere
- E. coli für Bakterien
- Bäckerhefe für Pilze
- Arabidopsis thaliana für Pflanzen

## Human Genome Project

### Ergebnisse

- menschl. Gen größer + mehr & größere Introns als Invertebrate
- Protein codierende Region ≈ 2% des Genoms
- alternative Splicing
- 97% des Genoms gleich bei allen Menschen
- 50% Transposons
- Gene ungleichmäßig auf Genom verteilt
  - Chr. 4: 231 Gene
  - Chr. 1: 2968 Gene

## Sequenzierung

Hybrid - Approach aus

- Hierarchical Shotgun Sequencing (HGP, Collins)
- Shotgun Sequencing (Celera, Craig)

## HGP

- Genomische DNA fragmentieren
  - Gen-library mit BAC
  - jeder Klon einzeln sequenziert

≈ Genom = viele Bücher  
→ jede Seite jedes Buchs einzeln sequenziert



- Reihenfolge bekannt



- viele Gaps

## Shotgun Seq.

- ohne BAC - library

≈ alle Wörter aller Bücher gleichzeitig sequenziert

⊖ Reihenfolge unbekannt + viele repetitive Sequenzen im menschl. Genom

## Hybrid

· Kombination beider Methoden:

Sequenzierte Reads überlappen

→ höhere Vollständigkeit

→ Assemblierung durch overlaps

→ Finishing: Gaps schließen → designs Primer mittels  
umliegender Region  
→ setzte an DNA an  
= Nachsequenzierung

## Coverage

= wie oft wurde eine Region sequenziert

≥ 10 × coverage für Genauigkeit von 1 Fehler / 10 000 Basen

## EST Sequencing

EST = expressed sequenced tags

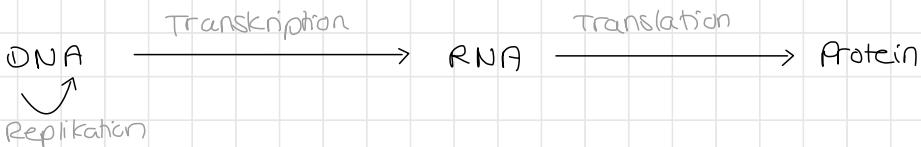
Idee: sequenziere nur "wichtigste" Gene

= Gene die in best. Gewebe exprimiert

→ Sequenziere cDNAs, die von RNA aus best. Gewebe hergestellt

# VIII. RNA & Genexpression

zentrales Dogma der MolBio:



## RNA

### Aufbau

- RNA = Polymer aus Monomeren
  - Monomer = Nukleotid
    - Phosphatgruppe
    - Pentosezucker (Ribose)
    - stickstoffhaltige Base (Cytosin / Guanin / Adenin / Uracil)
- bildet komplexe Sekundär-, Tertiärstrukturen aus

### Aufgaben

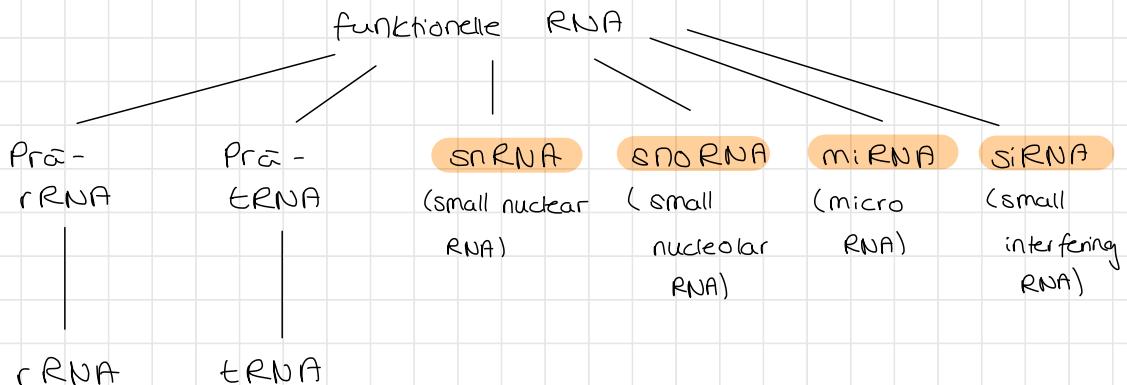
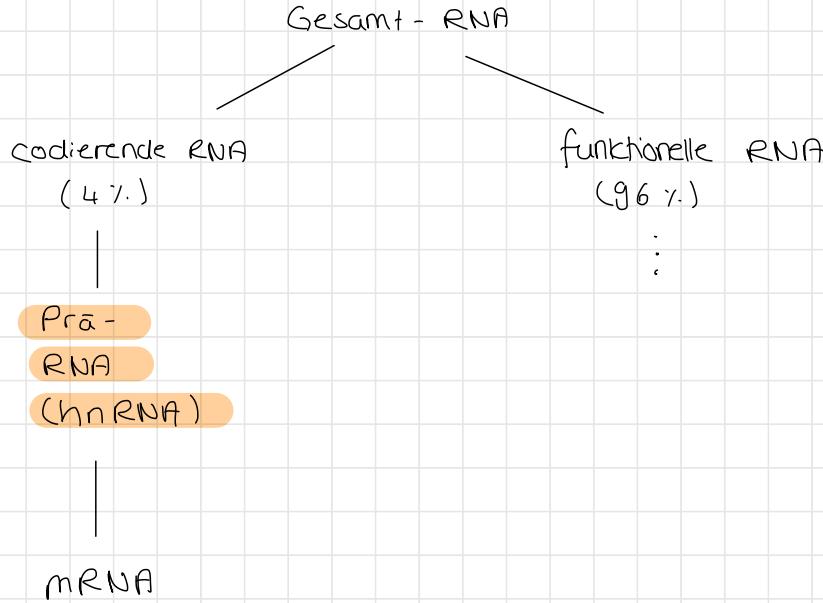
#### Beispiele

- Proteinbiosynthese
- Genexpression regulieren  
(rRNA, microRNA, Riboswitches)
- Peptidbindungen katalysieren  
(Ribozyme)

# RNA Arten

= nur bei

Eukaryoten



## Funktionen in Zelle

mRNA = messenger RNA

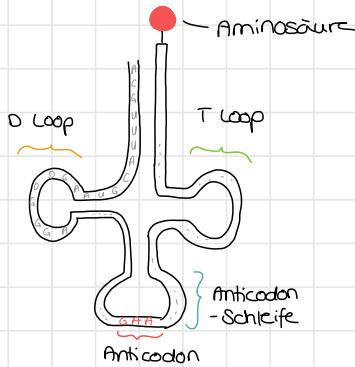
- trägt genetische Information, die von DNA kopiert wurde

rRNA = ribosomale RNA

- bilden mit ribosomalen Proteinen die Ribosomen

tRNA = transfer RNA

- eine tRNA je Aminosäure
- Schlüssel - Schloss - Prinzip: Anticodon zur mRNA



## Open Reading Frame (ORF)

- Bereich der DNA bzw. mRNA zwischen Start- & Stop-Codon
- = nicht-überlappende Folge von Codons
- codieren für AS-Sequenz eines Peptids / Proteins

Prokaryoten:

- ≥ 2 ORFs hintereinander
- codieren mehrere Peptidketten gleichzeitig
- = polycistronisch

- ORFs ohne Introns

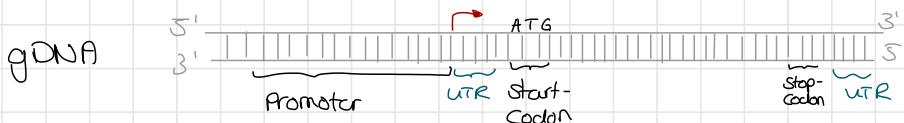
## Eukaryoten:

- immer nur 1 ORF
  - = monocistronisch
- ORFs mit Introns  
→ Splicing

## Proteinbiosynthese

### DNA → mRNA

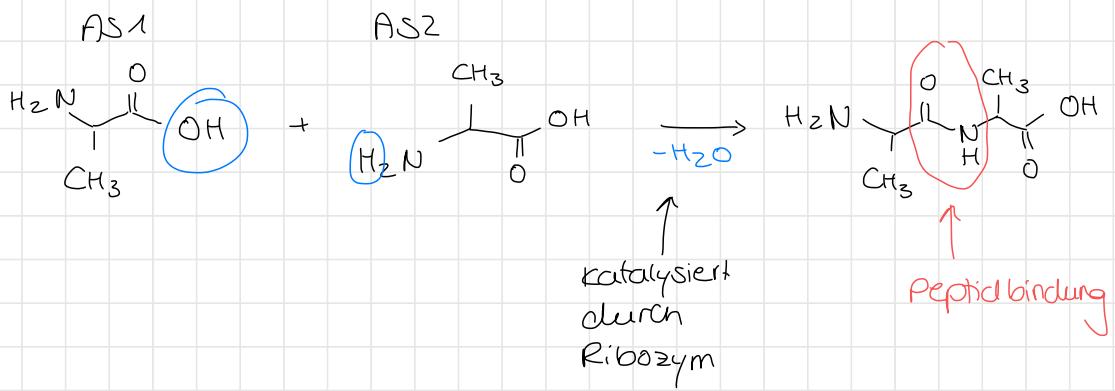
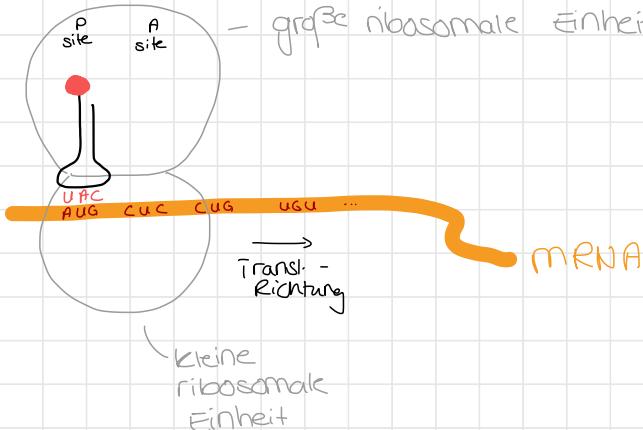
- Kopieren der genetischen DNA:
  - durch RNA-Polymerasen
  - Regulation u.a. über Promoter-Region



UTR = untranskribierte Region (Regulatoren)

- RNAp kopiert mit 3' - 5' Strang als Matrize
- ggf. Splicing
- Ribosom scannt mRNA, startet Synthese am Start-Codon (AUG)
  - mRNA = Vorlage
  - rRNA = bildet Ribosom (mit weiteren Proteinen)
  - tRNA = AS-Transporter
- Ribosom knüpft AS Kette bis Stop-Codon erreicht (UAA/ UAG/ UGA)

# Ribosom



## Transkription

- räuml. und zeitl. Variabel
  - alle Zellen gleiche DNA  
versch. mRNA
  - abh. vom Zelltyp, Zellzyklus

## Methoden

Zur Untersuchung der Genexpression

### Hybridisierungsbasiert

- Northern Blotting
- DNA Microarrays
- ...

### Sequenzierungsbasiert

- ESTs (Expressed Sequence Tags)
- RNA seq
- ...

### PCR - basiert

- RT - PCR (Reverse Transkriptase)
- Real Time PCR
- ...

## Northern Blotting

- ! Gel muss denaturierend sein
  - löse Sek. + Tertiärstrukturen auf

### Qualitätskontrolle

gesamte RNA einer Zelle in Gel → Gelenkstr. ph.

→ sollte 2 helle Banden zeigen, größere sollte dicker sein

= rRNA

= mit Abstand häufigste RNA in Zelle (kontin. Transkr.)

da mehr Nukleotide kann mehr mit tRNA interkalieren

→ 2 Banden = 28S & 18S - rRNA  
bzw. 23S & 16S - rRNA

### mRNA

- nur 1-5% der ges. RNA einer Zelle
- Visualisierung hybrid. mit Poly-T, Taq-Polymerase  
(bindet an mRNA Poly-A-Tail)

→ Problem: alle mRNAs versch. lang  
⇒ sehe nur Schmier in Gelelektr.ph.

### Forschungsfrage

- wie stark ist Gen X in versch. Zellen exprimiert?
1. Gelelekrophorese der ges. RNA in denaturierendem Gel
  2. Qualitätskontrolle über rRNA
    - gleiches Expressionsmuster in allen Zellen?  
z. B. Banden untersch. Dicke?
  3. RNA Transfer auf Nylon-Membran durch Blotting
  4. RNA auf Membran fixieren mit UV-Licht
  5. Hybridisieren der Membran mit denaturierter, markierter Probe
    - ↖ homolog zu ges. Gen X
  6. Visualisierung
    - je dicker die Bande von Gen X, desto höher die Expression in der Zelle (≈ relativ)
  7. Qualitätskontrolle mit Householding Gen
    - = gleiche Expression in allen Zellen  
z.B. rRNA
    - Aussagen nur möglich wenn Haus.h. Gen gleich exprimiert in allen Zellen

### Aussagen

- Größe + Menge der nachgewiesenen mRNA
- Expressionstarke (≈ nur geschätzt und relativ)
- Vergleich zw. Organen / Geweben möglich

- sehr viel Ziel- mRNA benötigt  
→ sehr aufwändig
- RNA sehr instabil

### Vergleich: Southern Blot

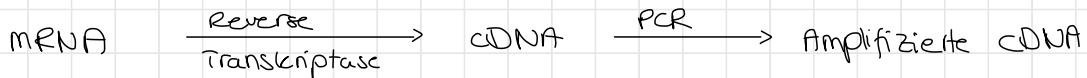
SB Aussagen:

- ob Gen vorhanden + wie viele Kopien im Genom
- Ähnlichkeit zw. chrom. Genen + Sorte
- Erkennungsseq. best. Restr.enzyme im Gen vorhanden
- Erstellen einer Restriktionskarte

? SB keine Aussage über Genexpression

### RT - PCR

= Reverse - Transcription PCR



### 1. sscDNA Synthese

- mRNA mittels Rev. Transkr. in single-stranded cDNA überführen

→ Rev. Transkr. braucht Primer

### 3 Mögliche Primer-Arten:

- Oligo dT - Primer: bindet an Poly-A der mRNA  
→ trifft unspezifisch alle mRNAs einer Zelle
- sequenzspez. Primer
- random Primer: bindet unspez. an zufällige Stellen der RNA

=> bindet an zufällige Stellen in mRNA  
= nur Teile des Gens als cDNA kopiert  
(partielle cDNA)

## Reverse Transkriptase

2 Enzymaktivitäten

- DNA Polymerase
- RNase H

~ zerstört RNA in RNA-DNA Hybriden ↴

=> brauchen Rev.Tr. mit ausgeschalteter RNase H

## 2. PCR Amplifizierung

der ssDNA mittels Taq Polymerase

-> brauche dNTPs + seq. spez. Primer

① nur semi-quantitativ

→ am Ende der PCR im Gel nur 10x - 100x Unterschiede erkennbar

## Real Time PCR (qPCR)

Idee: beobachte PCR während Amplifizierung

→ nutze SYBR-Green als Marker

- interkaliert mit dsDNA während PCR

-> bei Bindung ändert sich Emmissionsspektrum

## Methode:

- messe SYBR-Green Fluoreszenz nach jedem PCR Cycle + Plotten
- während PCR-Exponentialphase wird nach Ct Zyklen ein Threshold überschritten
  - Threshold = Fluoreszenzintensität deutlich über Hintergrundfluoreszenz des ungebundenen SYBR

- => Ct charakteristisch für Probe  
negativ korreliert mit cDNA Menge
- => je früher Ct erreicht desto mehr cDNA in Probe

## Quantifizierung

### Relativ

- Ct von Zielen verglichen mit Ct von Kontrollen  
(zB Haushaltsgen)

Anwendung:

- quantifizieren von Genexpression
- validieren von Array-Daten

### $\Delta\Delta Ct$ -Methode

- Target Gen = Gene of interest
- Referenz Gen = housekeeping Gene

Vergleich zweier Proben: behandelt vs. unbehandelt  
zB Tumor gesund

$$\Delta Ct_{untreated} = Ct_{target; u} - Ct_{reference; u}$$

$$\Delta Ct_{treated} = Ct_{target; t} - Ct_{reference; t}$$

$$\Delta\Delta Ct = \Delta Ct_{untreated} - \Delta Ct_{treated}$$

$$\text{Fold change} = 2^{\Delta\Delta Ct}$$

= wie hoch ist Unterschied der Genexpr.  
zw. treated, untreated

## -ΔCt - Methode (P Minus)

$$\Delta Ct_{untreated} = Ct_{target; u} - Ct_{reference; u}$$

$$\Delta Ct_{treated} = Ct_{target; t} - Ct_{reference; t}$$

$$\Delta\Delta Ct = \Delta Ct_{treated} - \Delta Ct_{untreated}$$

(umgekehrt zu oben)

$$\text{Fold change} = 2^{-\Delta\Delta Ct}$$

## Absolut

- = bestimmen der exakten # Genkopien
- vergleich mit Standardkurve

- Anwendung:
- Qualitätskontrolle
  - klinische Diagnosen
  - Mutationsdetektion
  - Lebensmittelkontrollen

## Standardkurve

x - Achse = Starting Quantity = # Kopien in Probe

y - Achse = Ct

=> schaue nach x - Wert zum Ct - Wert

## Genomweite Analyse

Ziel: welche Gene sind wann und wo exprimiert

Schritt 1: cDNA Genbanken erstellen

### SMART - Methode

### differenzielle Methode

zur Herstellung von cDNA

1. mRNA mit PolyA-Tail = Startmaterial
  2. Oligo dT Primer bindet an Poly A  
↳ hat direkt PCR-Primer-Adapter mit dabei
  3. Reverse Transkr. synthet. Gegenstrang  
→ am Ende baut RT Poly-C ein
  4. Adapter mit Poly-G bindet an Poly-C  
↳ hat 2. PCR-Primer dabei
  5. DNA-RNA Hybrid wird abgebaut
  6. DNA-Matrix wird komplementiert  
= ds cDNA
  7. Restriktionsverdau + Einbau in Vektor  
⇒ cDNA Library
- für jeden Zelltyp Library der RNA erstellen  
→ cDNA je Zelltyp mit versch. Farben markieren

zB      Neuron  
cDNA  
grün gelabelt

Epithel  
cDNA  
rot gelabelt

⇒ Hybridisierung mit cDNA aus Neuron

⇒ falls Hybridisierung erfolgreich: Farbe detektierbar

· bei Neuron sollte alles leuchten = Kontrolle

· bei Epithel leuchten nur die Gene die sowohl in Neuron als auch in Epithel exprimiert sind

⇒ Detektion von Genen die nur in best. Zellen benutzt

### Microarrays

warum kein North. Blot? Genanzahl die auf einmal prüfbar limitiert bei # Spuren im Gel

→ Microarray: alle Gene eines Genoms parallel

= cDNA Library

1. alle Gene auf Microarray fixiert = Probe  
→ 1 Punkt / Gen      ≈ Glas, Plastik, Nylon

2. mRNA / cDNA ist markiert = Target  
≈ zB C<sub>y3</sub> = grün, C<sub>y5</sub> = rot

3. Hybridisierung:  
- cDNAs aus 2 versch. Zellen auf Microarray  
zB gesund vs. Tumor

4. Detektion:

- je nach Farbe entscheiden in welcher Zelle welches Gen exprimiert

(Keine Farbe / 1 Farbe / Mischfarbe?)

=> Aussagen:

- welches Gen in welcher Probe exprimiert
- relativer Vergleich der Starke zw. 2 Proben

### RNA - Seq

Transkriptom = alle transkribierten RNA

in einer Zelle

- Sequenzierungs basiert
- Transkriptom Shotgun - Sequencing  
→ assemblieren an Referenz - Genom

1. RNA Isolation aus Zelle
2. Qualitätskontrolle (wie bei North. Blot)
3. Waschen mit DNase → degradiert DNA
4. RNA Fragmentierung + cDNA Synthese
5. Selektion nach Fragmentgröße  
(! snRNA verloren)
6. Anhängen von Seq. adaptoren + Sequenzierung
7. Assemblierung an Referenzgenom

### Probleme

- RNA = kurze Exons durch lange Introns getrennt  
→ Assembl. schwierig
- RNA Abundanz variiert stark  
→ rRNA sehr häufig
- RNA versch. lang