

Historie a metody genomiky



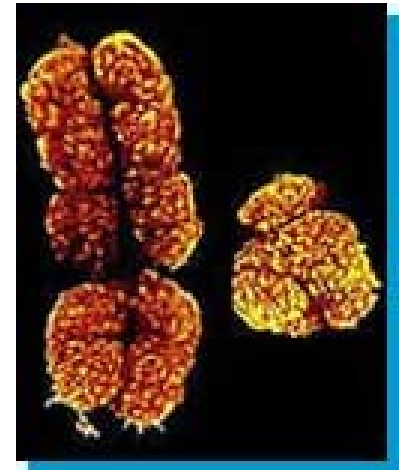
Eduard Kejnovský a Roman Hobza

Objev pohlavních chromosomů

**1891 Henking pracoval na spermatogenezi
u *Pyrrhocoris apterus***

**Polovina dceřiných buněk měla jeden
element (chromosom) navíc**

X bylo označení neurčitosti pozorovaného

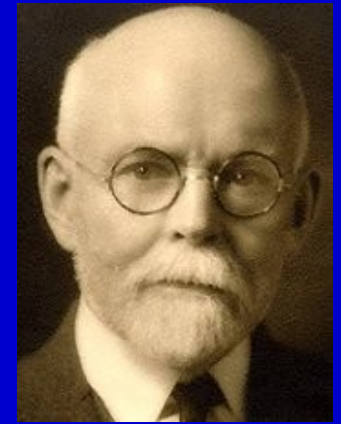


**1902 McClung označil tento element jako „accessory“
chromosom (X chromosome)**



Nettie Stevens

Chromosomální determinace pohlaví (1905)

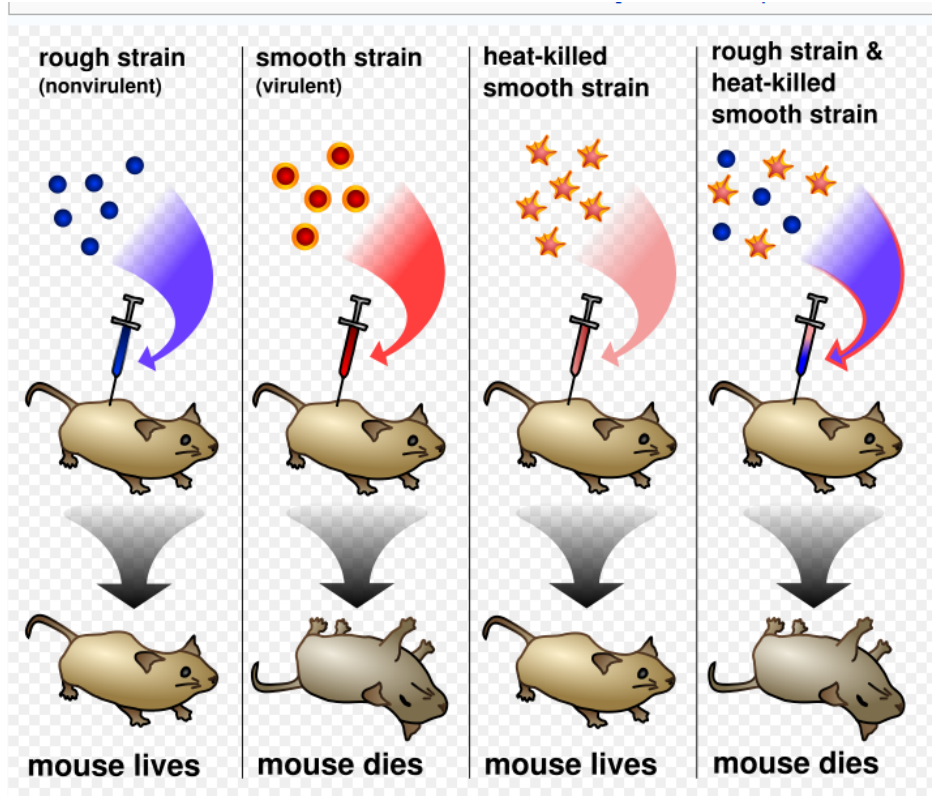


Edmund Wilson

Tenebrio molitor, *Coleoptera*
(F = AAXX, M = AAXY)
mealworm



Přeměna neškodných bakterií ve virulentní



1928 - Fred Griffith: experimenty na *Streptococcus pneumonia*

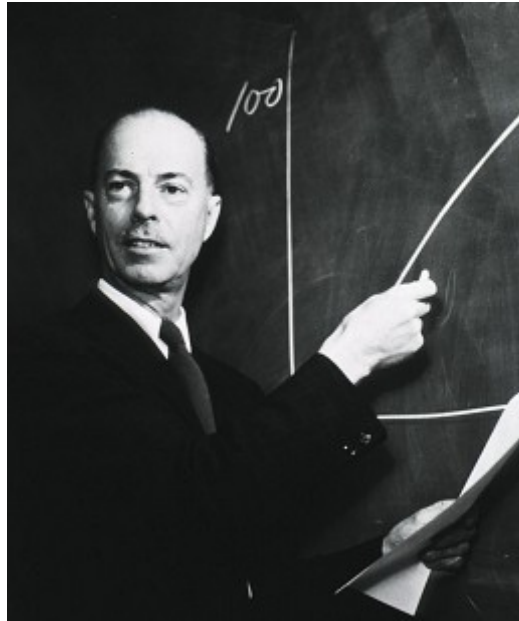
Dva kmeny:

Smooth (S) - virulentní

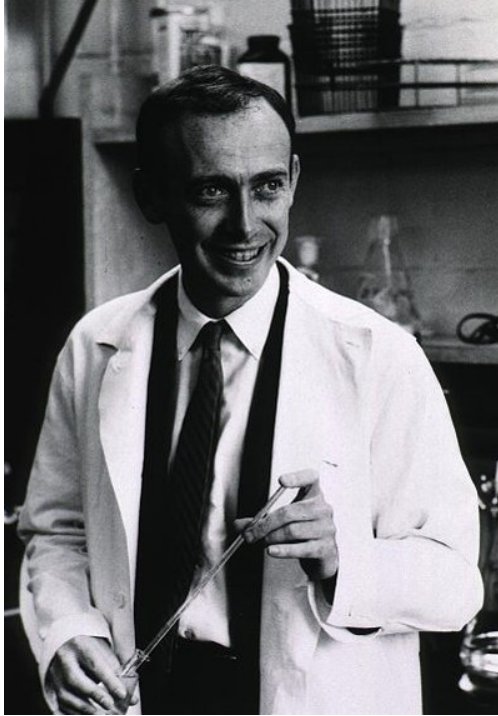
Rough (R) - nevirulentní

Injeckce R a teplotně inaktivovaných S bakterií - myši zemřely a obsahovaly S bakterie



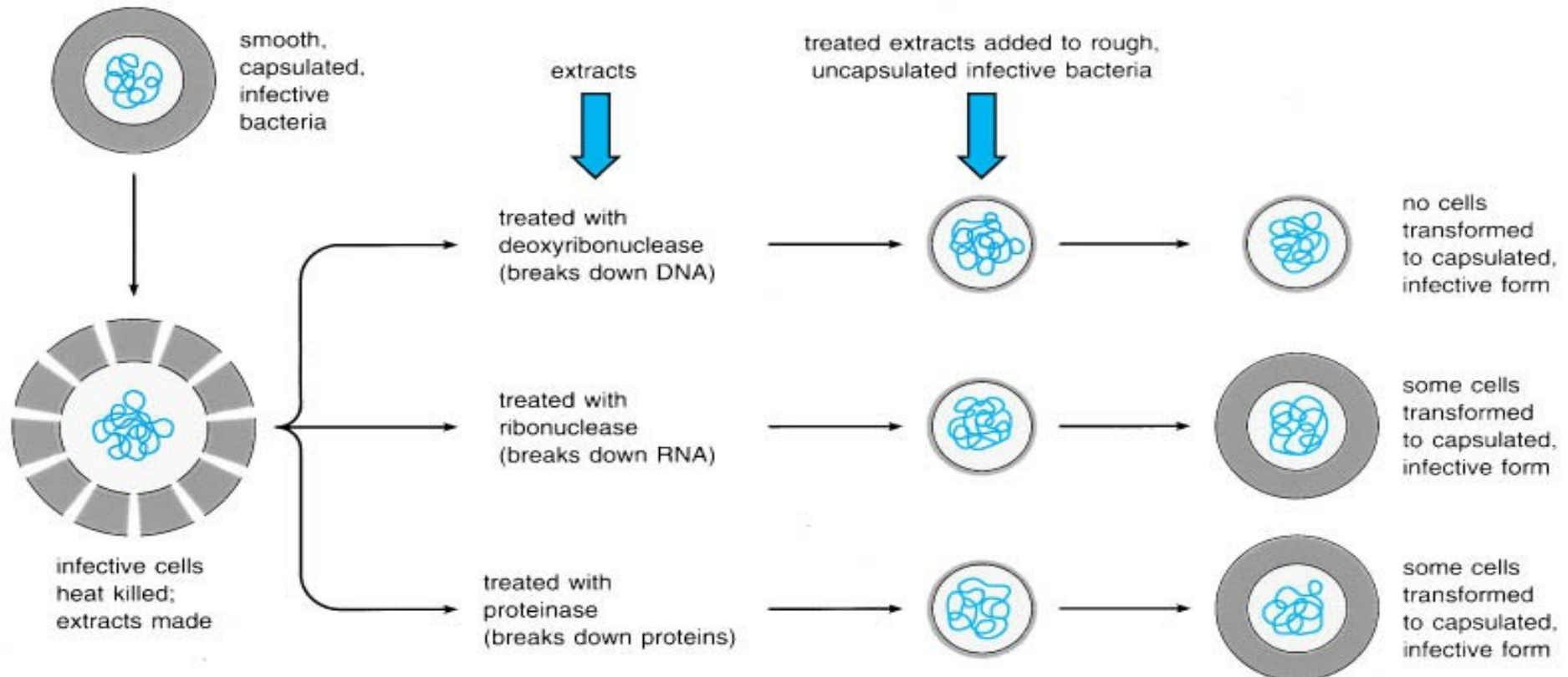






1944 - Avery, MacLeod, and McCarty

- Smíchali R kmen s DNA z S kmene a izolovali S bakterie
- Přidali DNase - inhibice transformace
- Přidali proteasy - transformace proběhla
- DNA určena jako “transformační princip”



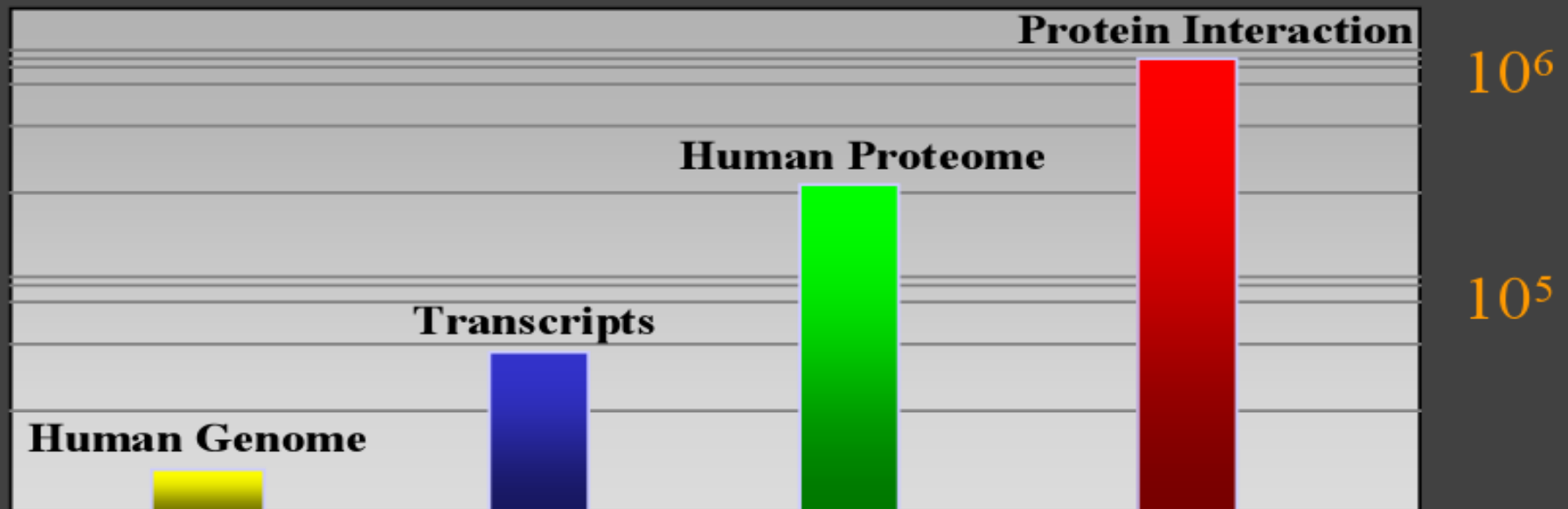
Genome: 30.000 genes

↓

Transcriptome: 40-100.000 mRNAs

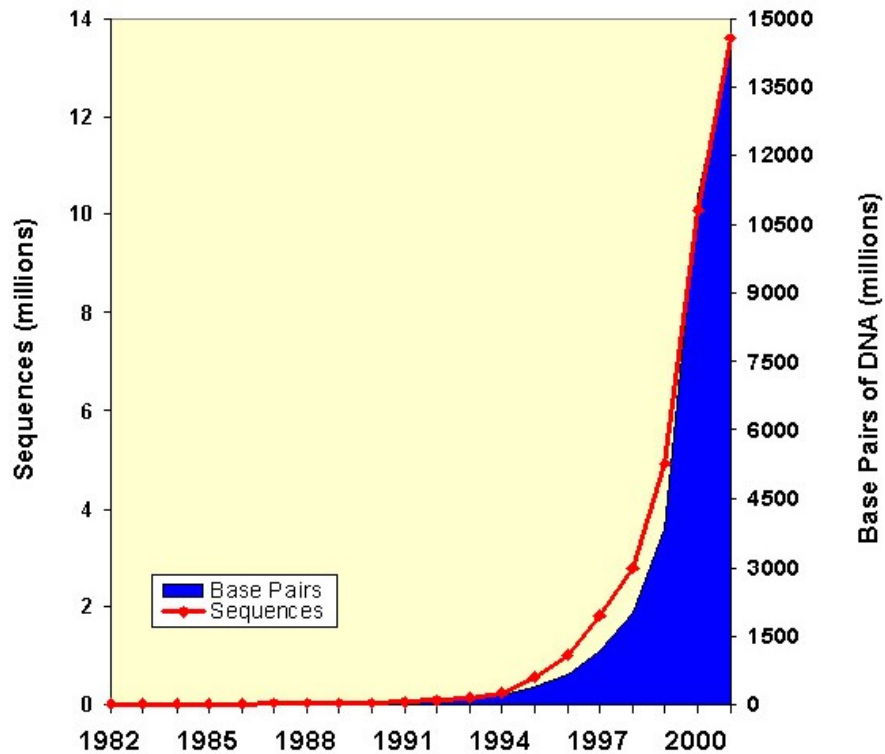
↓

Proteome: 100-400.000 proteins
>1.000.000 interactions



Sekvenování genomů

- 1988 americký kongres stál u zrodu Genbank





Applied Biosystems
ABI 3730XL
1 Mb / day



Roche / 454
Genome Sequencer FLX
100 Mb / run



Illumina / Solexa
Genetic Analyzer
2000 Mb / run



Applied Biosystems
SOLiD
3000 Mb / run



PACIFIC
BIOSCIENCES®



ion torrent
◊ * △ ○ × □ + ∞



Oxford
NANOPORE
Technologies®



illumina inc. (ILMN) - NasdaqGS

[Add to Portfolio](#)

[Like](#) 17

48.51 +0.07 (0.14%) 11:41AM EST - Nasdaq Real Time Price

Enter name(s) or symbol(s) [GET CHART](#) [COMPARE](#) [EVENTS](#) [TECHNICAL INDICATORS](#) [CHART SETTINGS](#) [RESET](#)



Pacific Biosciences of Californ

■ PACB

Nov 14, 2012





Comparison to other sequencing methods

	Ion Torrent ^{[14][16]}	454 Sequencing ^[17]	Illumina ^[18]	SOLiD ^[19]
Sequencing Chemistry	Ion semiconductor sequencing	Pyrosequencing	Polymerase-based sequence-by-synthesis	Ligation-based sequencing
Amplification approach	Emulsion PCR	Emulsion PCR	Bridge amplification	Emulsion PCR
Mb per run	100	100	600,000	170,000
Time per run	1.5 hours	7 hours	9 days	9 days
Read length	200 bp	400 bp	2x150 bp	35x75 bp
Cost per run	\$ 350 USD	\$ 8,438 USD	\$ 20,000 USD	\$ 4,000 USD
Cost per Mb	\$ 5.00 USD	\$ 84.39 USD	\$ 0.03 USD	\$ 0.04 USD
Cost per instrument	\$ 50,000 USD	\$ 500,000 USD	\$ 600,000 USD	\$ 595,000 USD

Proč sekvenovat dál?

- Komparativní genomika
- Biomedicínský výzkum
- Osobní genom

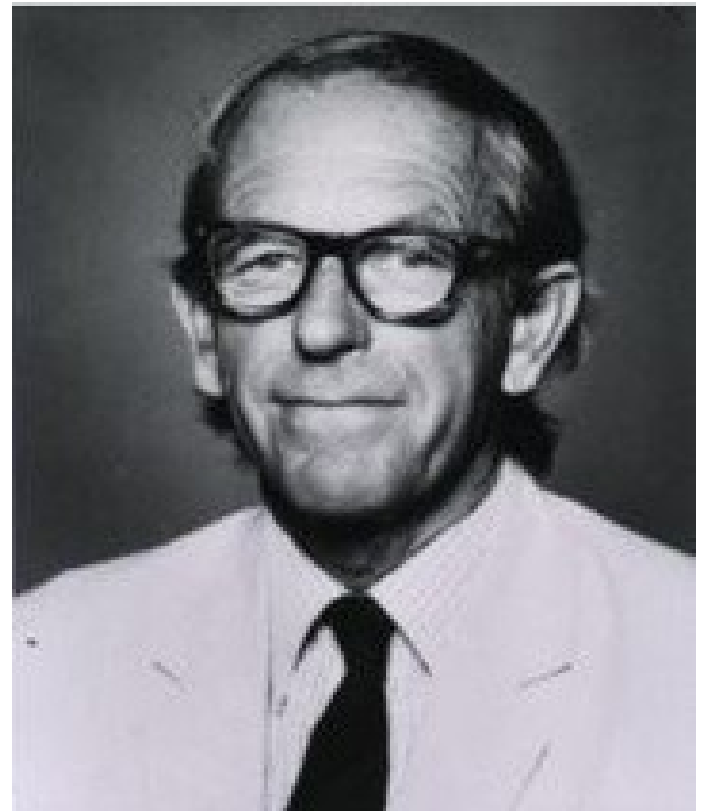
Frederick Sanger

1975 - Dideoxy sekvenační metoda

1977 – osekvenoval Φ -X174 (5,368 bp)

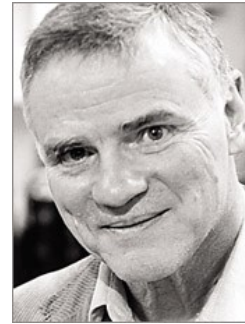
**1980 – dostal druhou Nobelovu cenu
za chemii**

**Později (polovina 80-tých let)
osekvenoval bakteriofága λ pomocí
shotgun metody (48,502 bp)**



Sekvenování genomů

- **1986** Leroy Hood: první automatický sekvenátor
- **1986** Human Genome Initiative
- **1990** započat projekt sekvenování lidského genomu (předpokládaná doba 15 let)

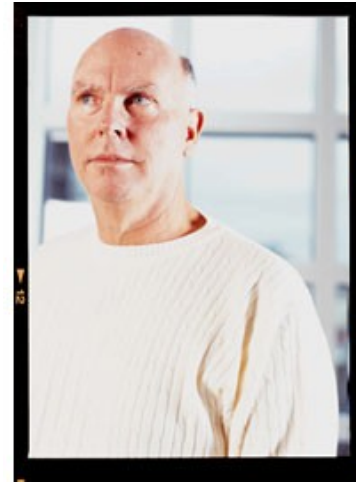


Leroy Hood



Sekvenování genomů

- **1995** John Craig Venter sekvenoval první bakteriální genom
- **1996** první eukaryotický genom (kvasinka) sekvenován



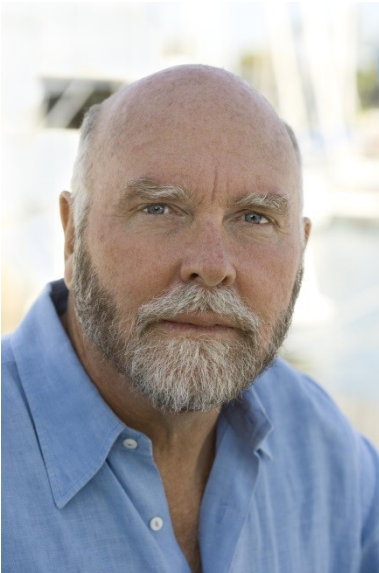
John Craig Venter

Craig Venter

Global Ocean Sampling Expedition

Synthetic genomics

Human Longevity Inc



<http://www.youtube.com/watch?v=J0rDFbrhjtI>

Sekvenování genomů

- **1997** sekvence *E. coli*
- **1998** *Caenorhabditis elegans* genom (první multicekulární genom)
- **1999** lidský chromozom 22 sekvenován

2010 Ideální lidský genom sekvenován

2010 Ideální lidský genom sekvenován



Sekvenování genomů

- **2000** *Drosophila melanogaster* genom
- **2001** Human Genome Sequencing:
předběžná sekvence lidského genomu

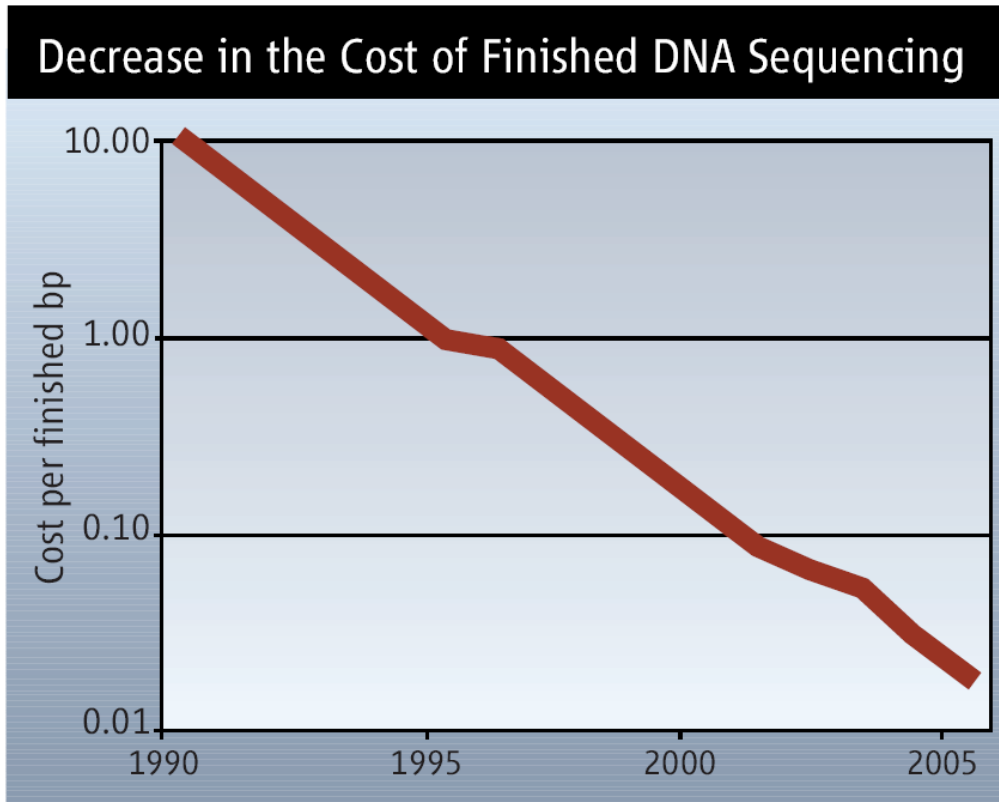


Sekvenování genomů

- **duben 2003** Lidský genom.
Sekvence myšího genomu.
- **duben 2004**
Sekvence krysího genomu



Honba za sekvencemi



Ø Human genome (first draft) –
\$300 million (2001)

Ø Rhesus macaque –
\$22 million (2006)

Ø Mammalian genome –
\$100,000 (by the end of 2006)

Free fall. As with computer technology, the plunging cost of DNA sequencing has opened new applications in science and medicine.

The Race for the \$1000 Genome. *Science* 311: 1544 – 1546, 2006

Honba za sekvencemi

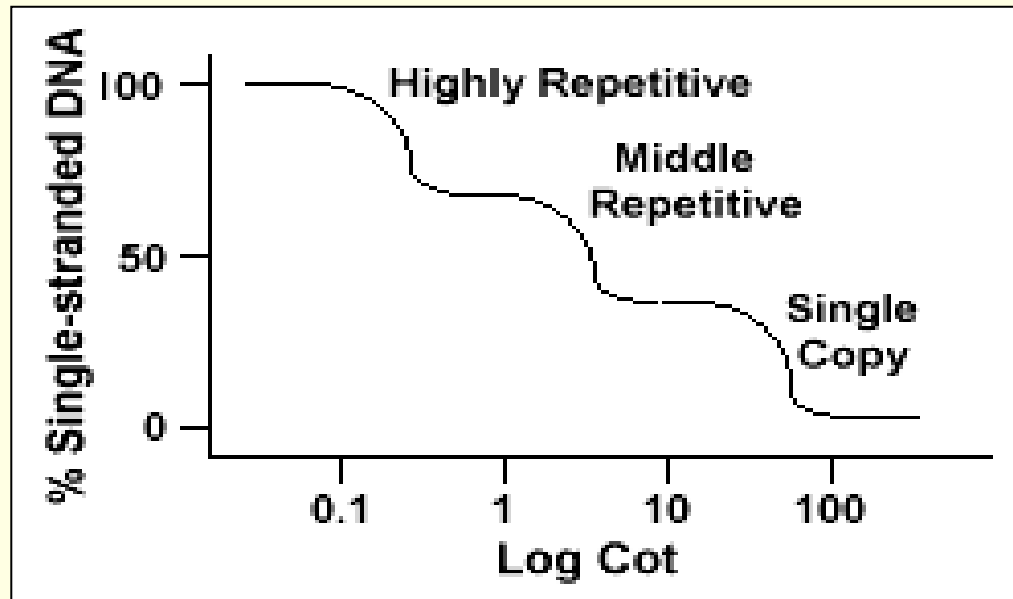
- X prize foundation

10,000,000\$ tomu, kdo osekvenuje během 10 dnů sto lidských genomů (Hawking.....)

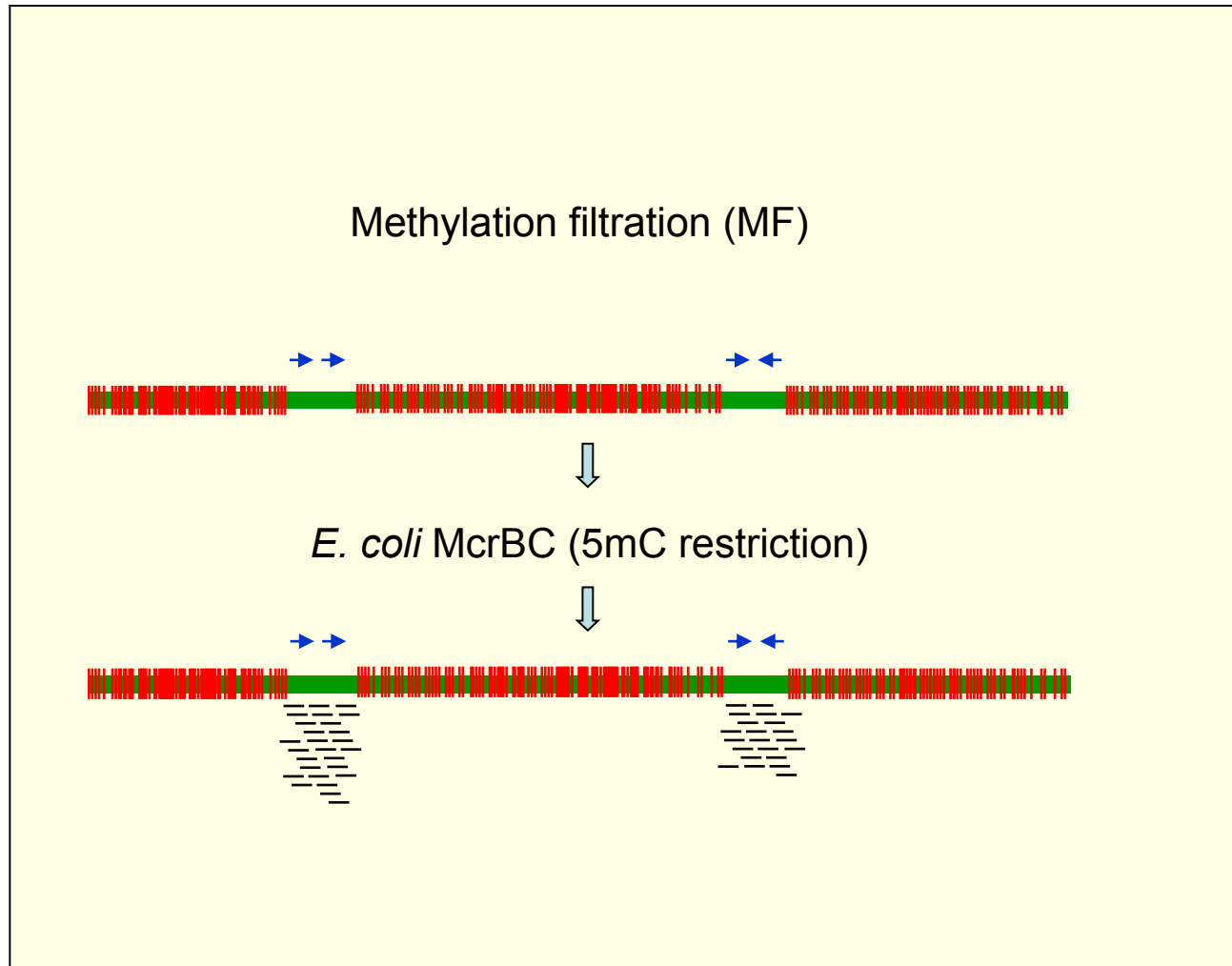


Redukce kopmlexity

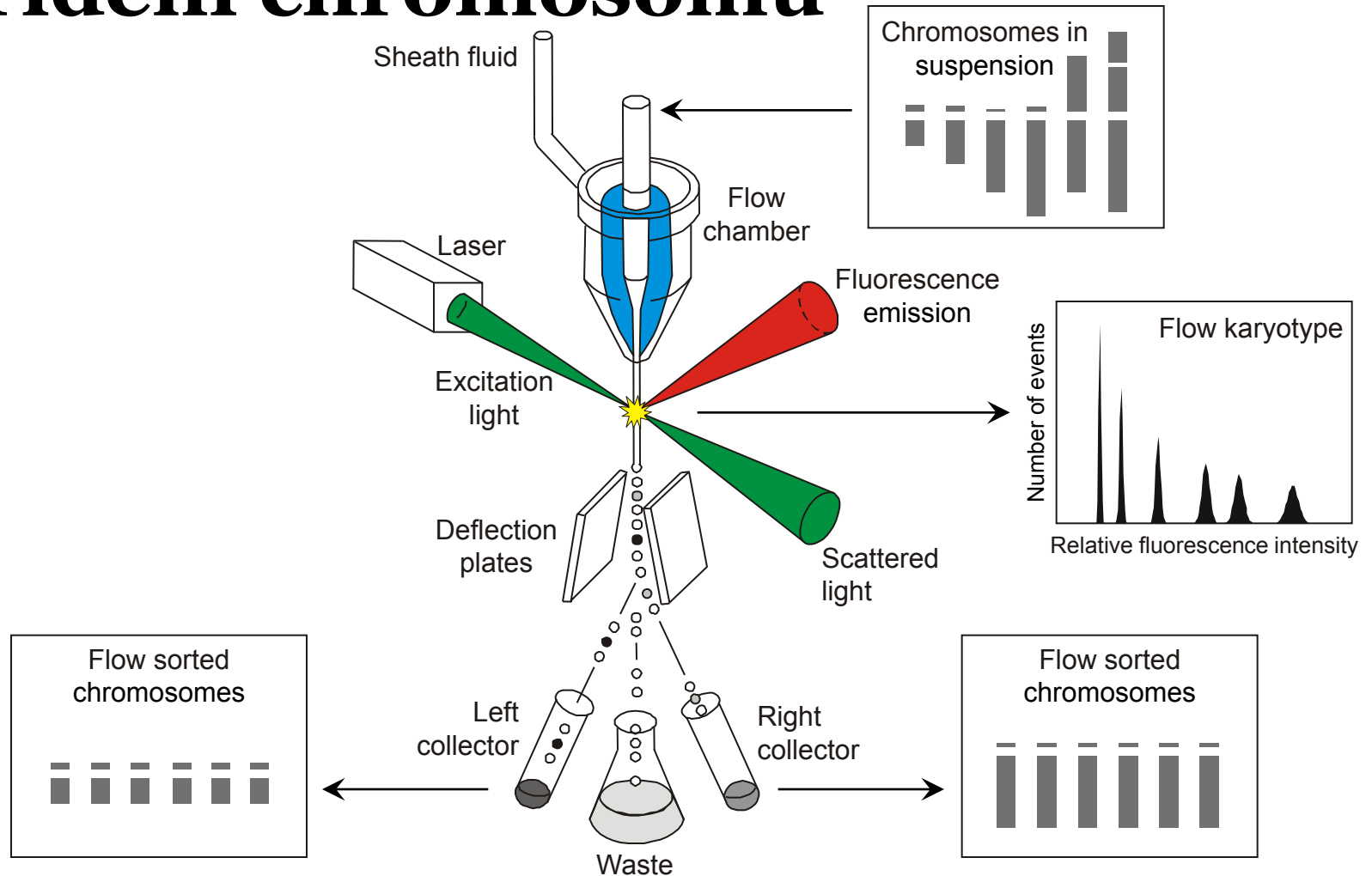
Hi-Cot selection



Redukce kopmlexity



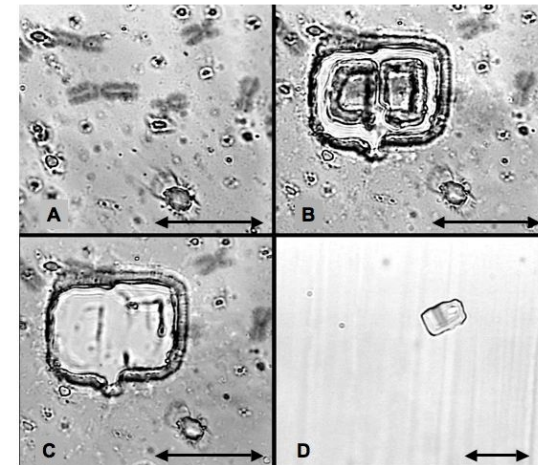
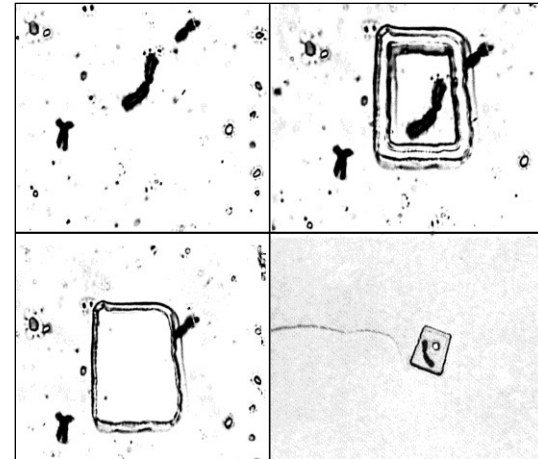
Redukce kopmlexity – třídění chromosomů



Laserová mikrodisekce

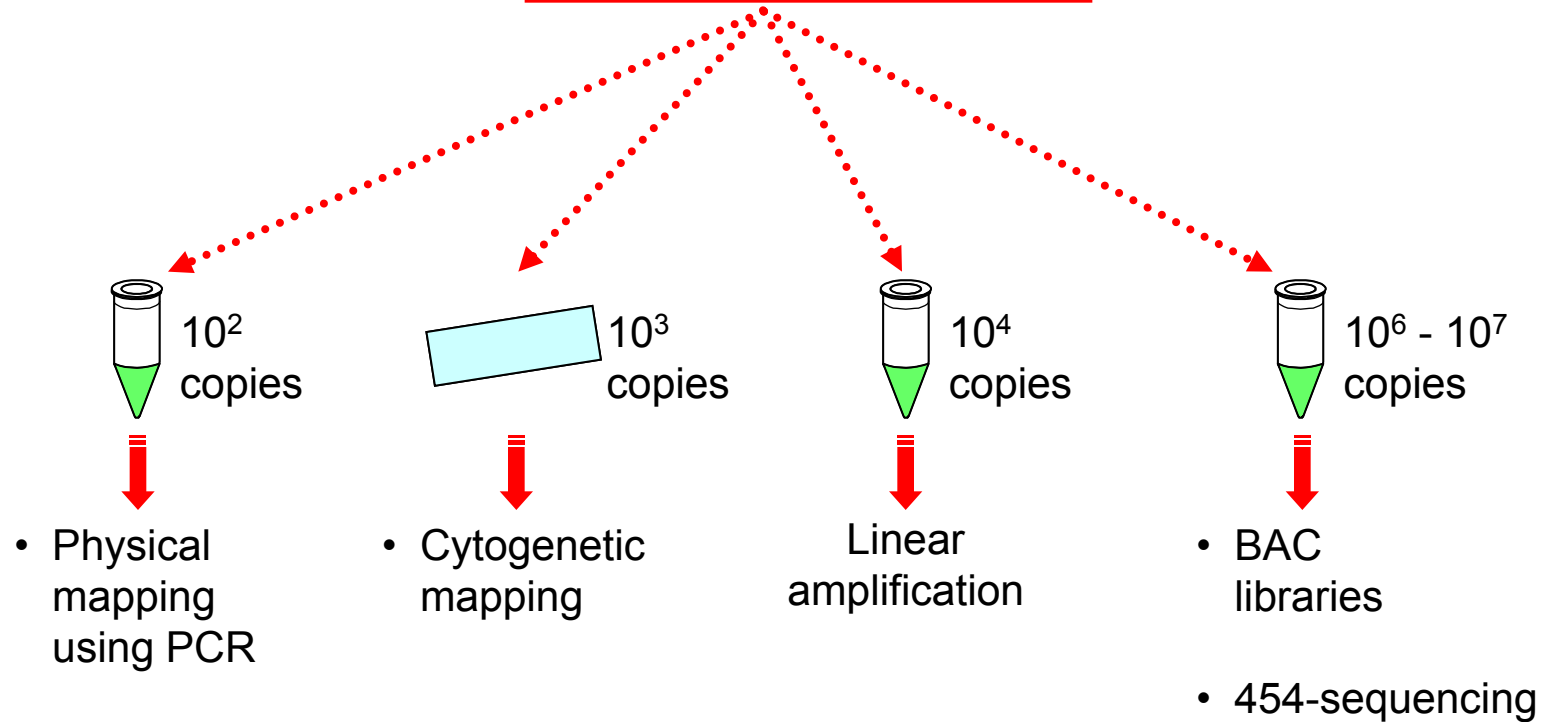
Výhody: vysoká čistota

Nevýhody: malý počet chromozomů, pracnost

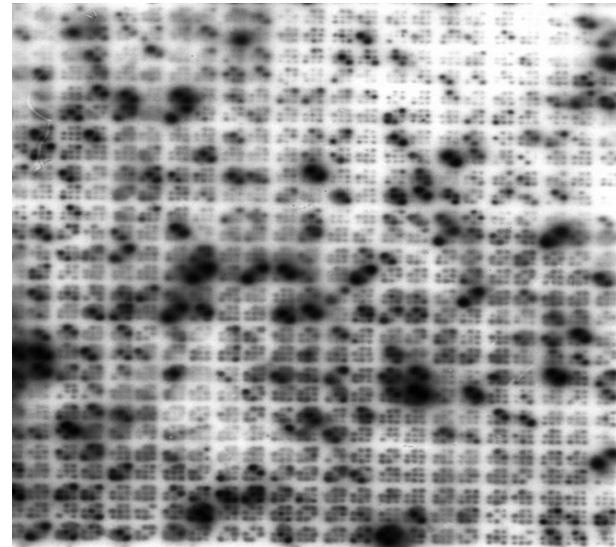
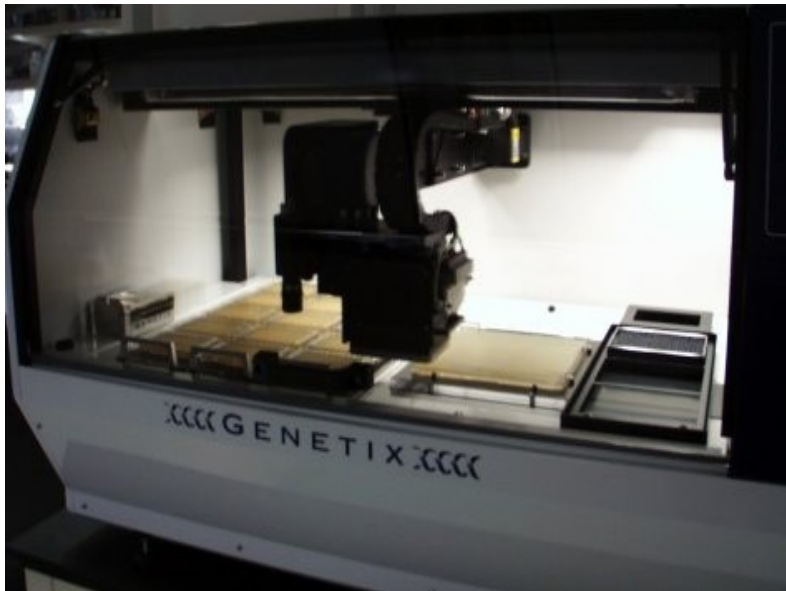


Aplikace

Separované chromosomy



Chromosomálně specifické genomové knihovny



Malování chromosomů – chromosome painting

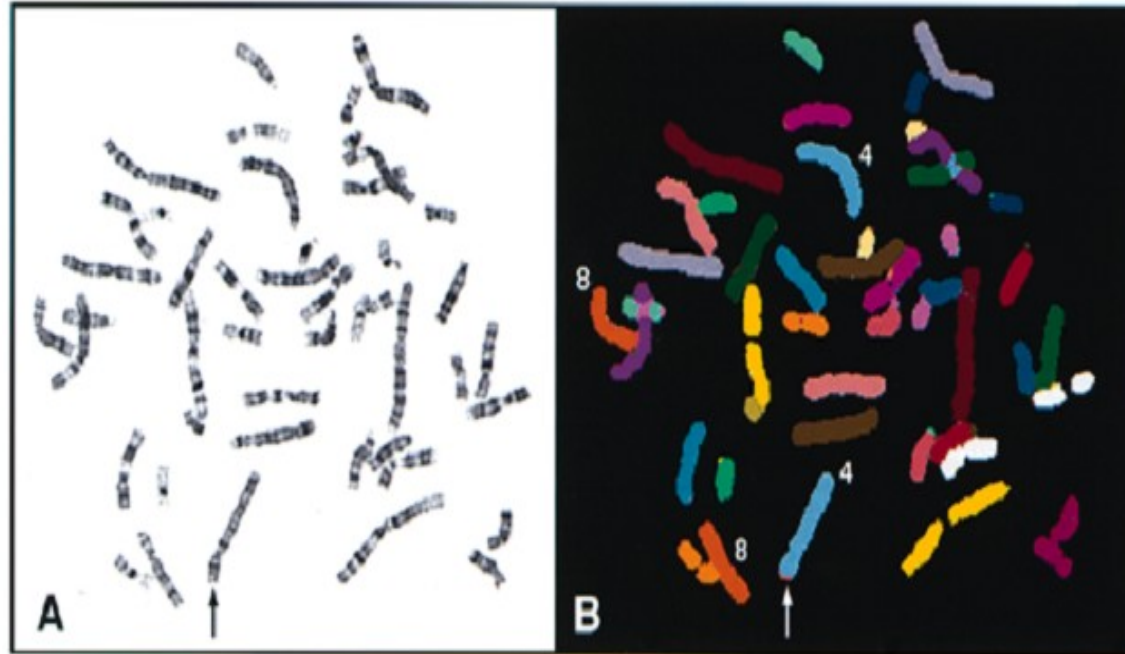


Figure 1. SKY analysis of a patient with Wolf–Hirschhorn syndrome. (A) G-banded metaphase. The telomeric region of one chromosome 4 (arrow) shows a slightly abnormal banding pattern. (B) The same metaphase as shown in (A) after SKY analysis. Chromosome classification revealed an unbalanced translocation between chromosomes 4 and 8 [46,XY,der(4)t(4;8)].



Applied Biosystems
ABI 3730XL
1 Mb / day



Roche / 454
Genome Sequencer FLX
100 Mb / run



Illumina / Solexa
Genetic Analyzer
2000 Mb / run



Applied Biosystems
SOLiD
3000 Mb / run

Metody sekvenování

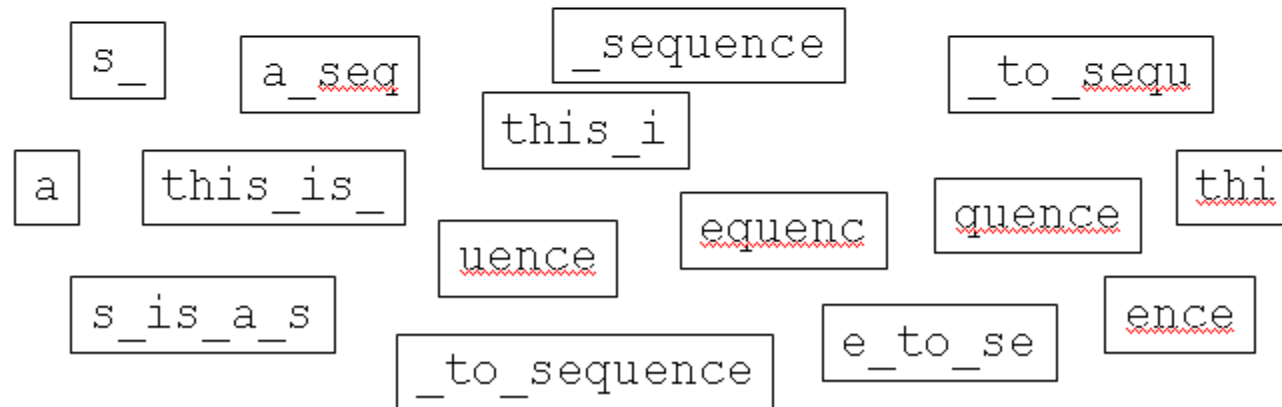
this_is_a_sequence_to_sequence

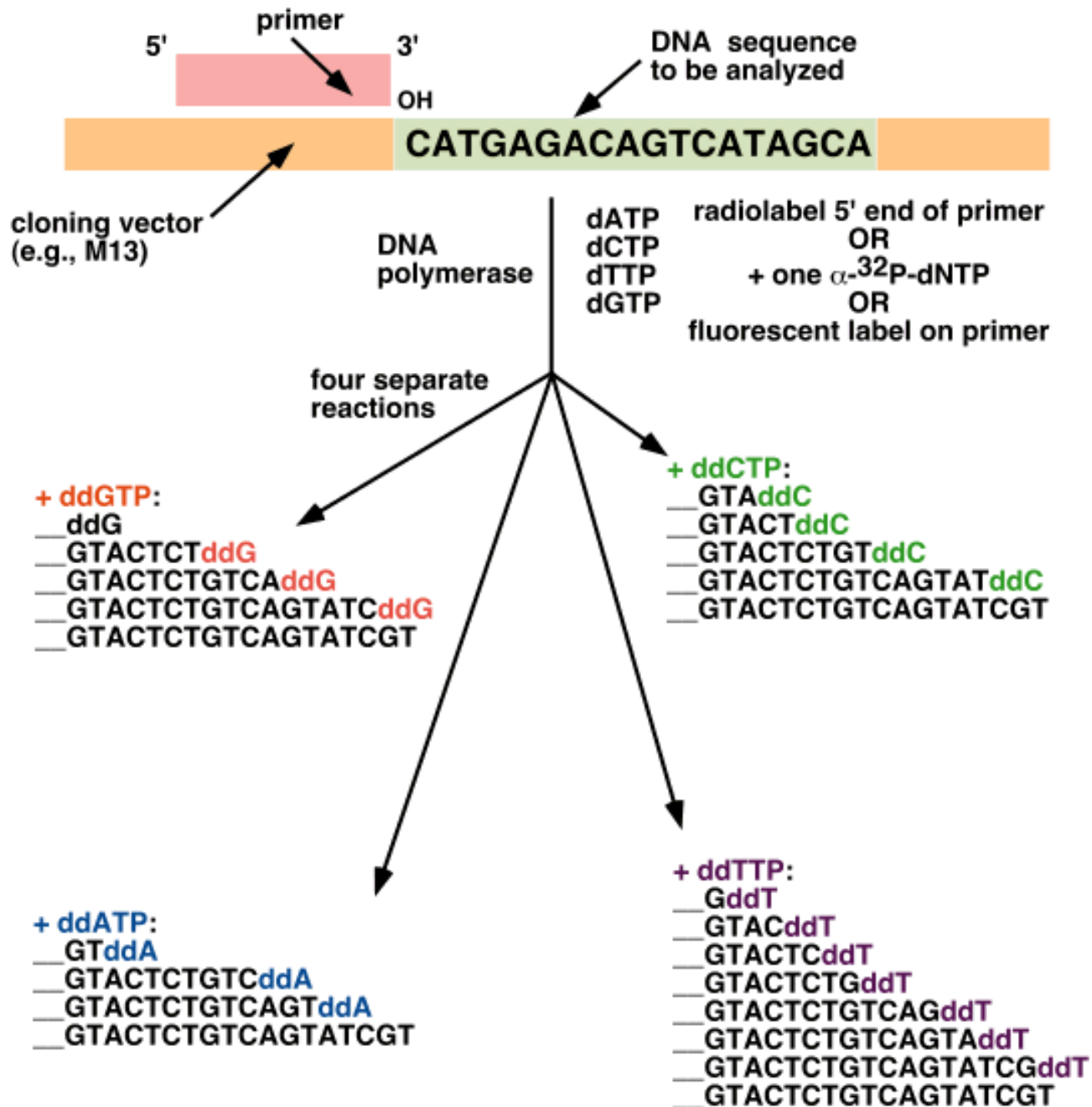


this_is_a_sequence_to_sequence

this_is_a_sequence_to_sequence

this_is_a_sequence_to_sequence





+ ddGTP:
 _ ddG
 _ GTACTCTddG
 _ GTACTCTGTCAddG
 _ GTACTCTGTCAGTATCddG
 _ GTACTCTGTCAGTATCGT

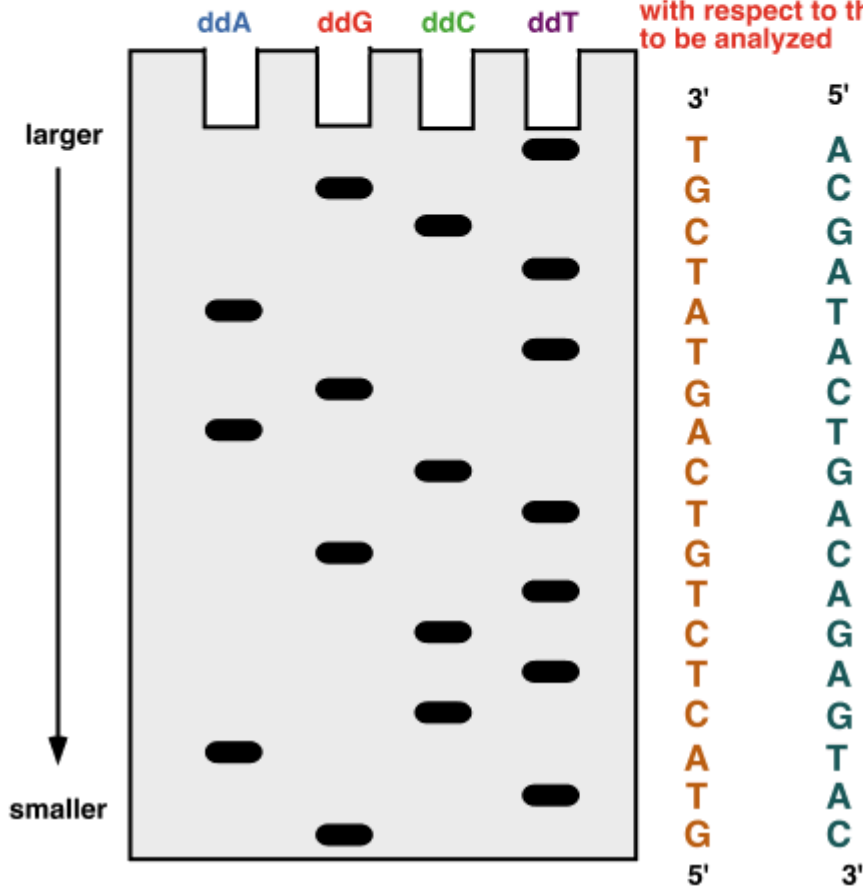
+ ddCTP:
 _ GTAddC
 _ GTACTddC
 _ GTACTCTGTddC
 _ GTACTCTGTCAGTATddC
 _ GTACTCTGTCAGTATCGT

+ ddATP:
 _ GTddA
 _ GTACTCTGTCddA
 _ GTACTCTGTCAGTddA
 _ GTACTCTGTCAGTATCGT

+ ddTTP:
 _ GddT
 _ GTACddT
 _ GTACTCddT
 _ GTACTCTGddT
 _ GTACTCTGTCAGddT
 _ GTACTCTGTCAGTAddT
 _ GTACTCTGTCAGTATCGddT
 _ GTACTCTGTCAGTATCGT

gel electrophoresis
 autoradiography (if radiolabeled)

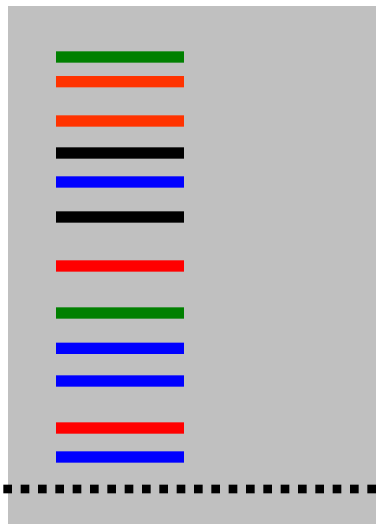
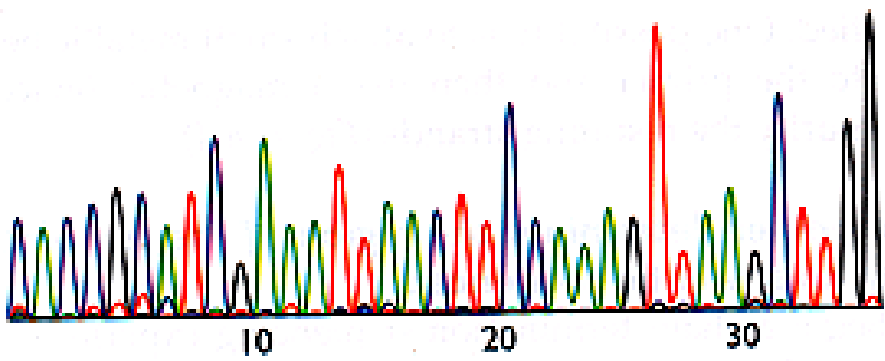
sequence read from gel
 is the complementary strand
 with respect to the sequence
 to be analyzed



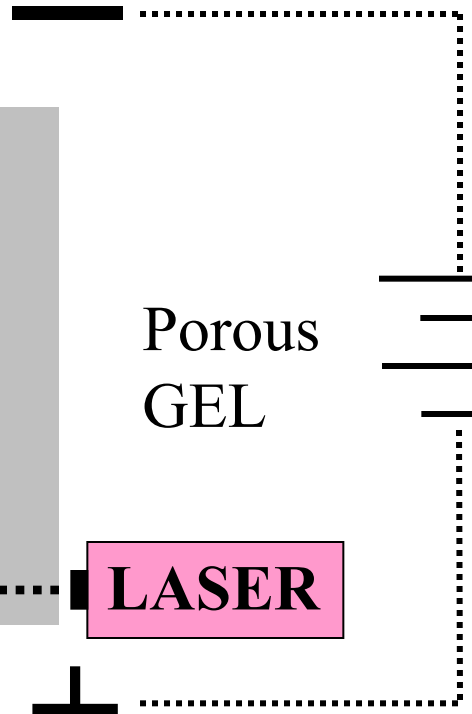
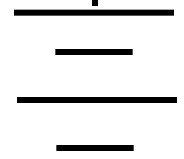
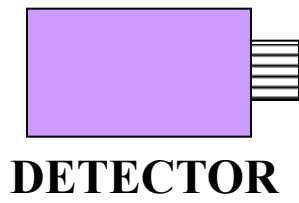
aggctcctctcccacc



a
ag
agg
aggc
aggct
aggctc
.....



Porous
GEL



Genome Sequencer 20 System 454 pyrosequencing (2005)

- <http://www.454.com>



Neanderthal
sequenced *now!*

Max Planck Institut
uses 454 Sequencing™
technology.

Read about it →

Watch the **CNN** video →

Hear the **NPR** broadcast →

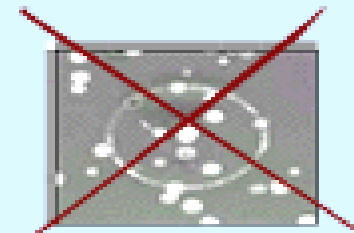
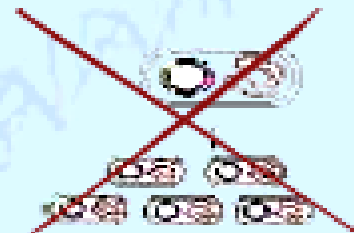
Image credit:
Ken Mowbray and Blaine Maley,
American Museum of Natural History

DNA library preparation

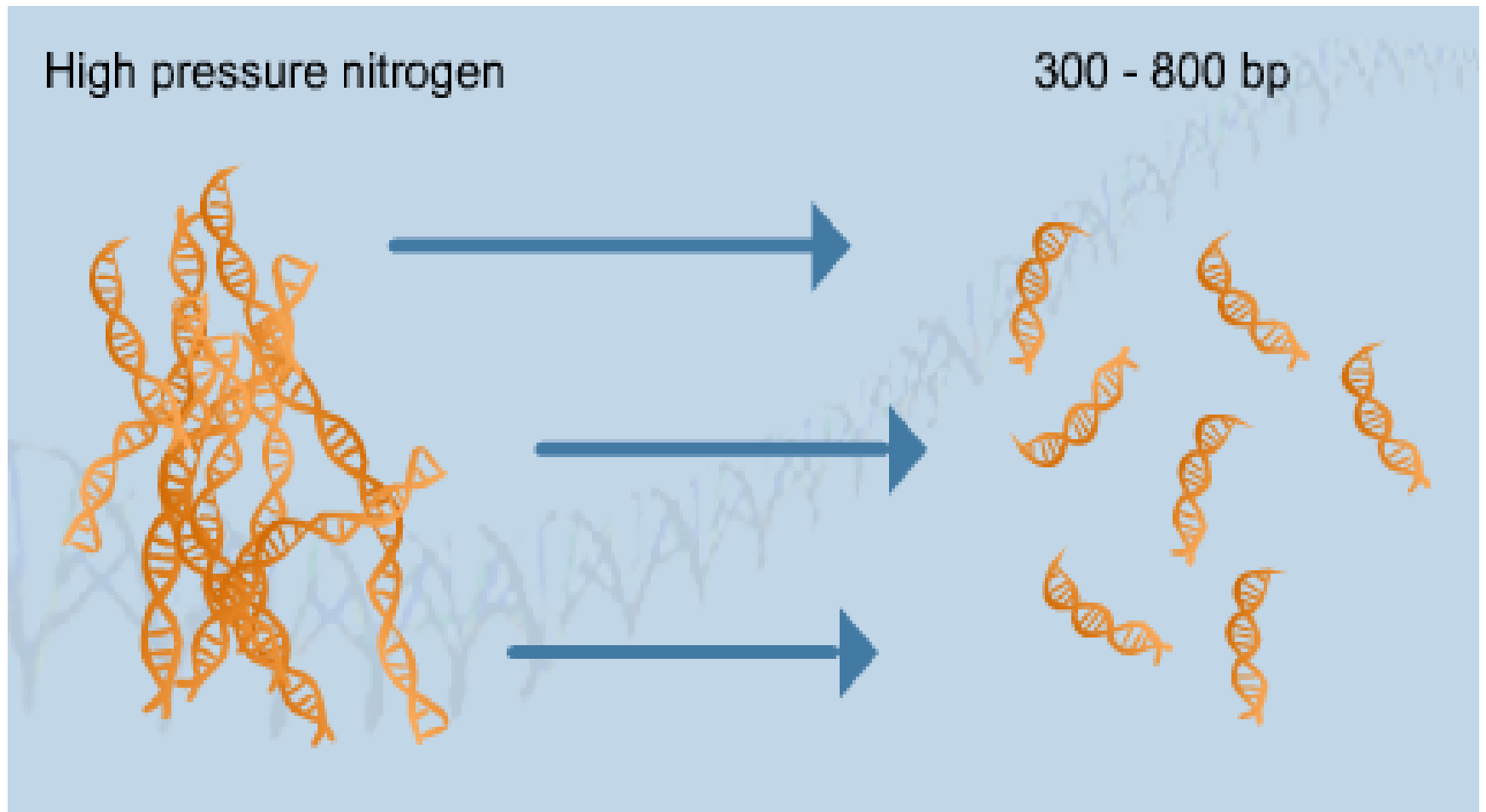
One sample preparation per genome

No Cloning

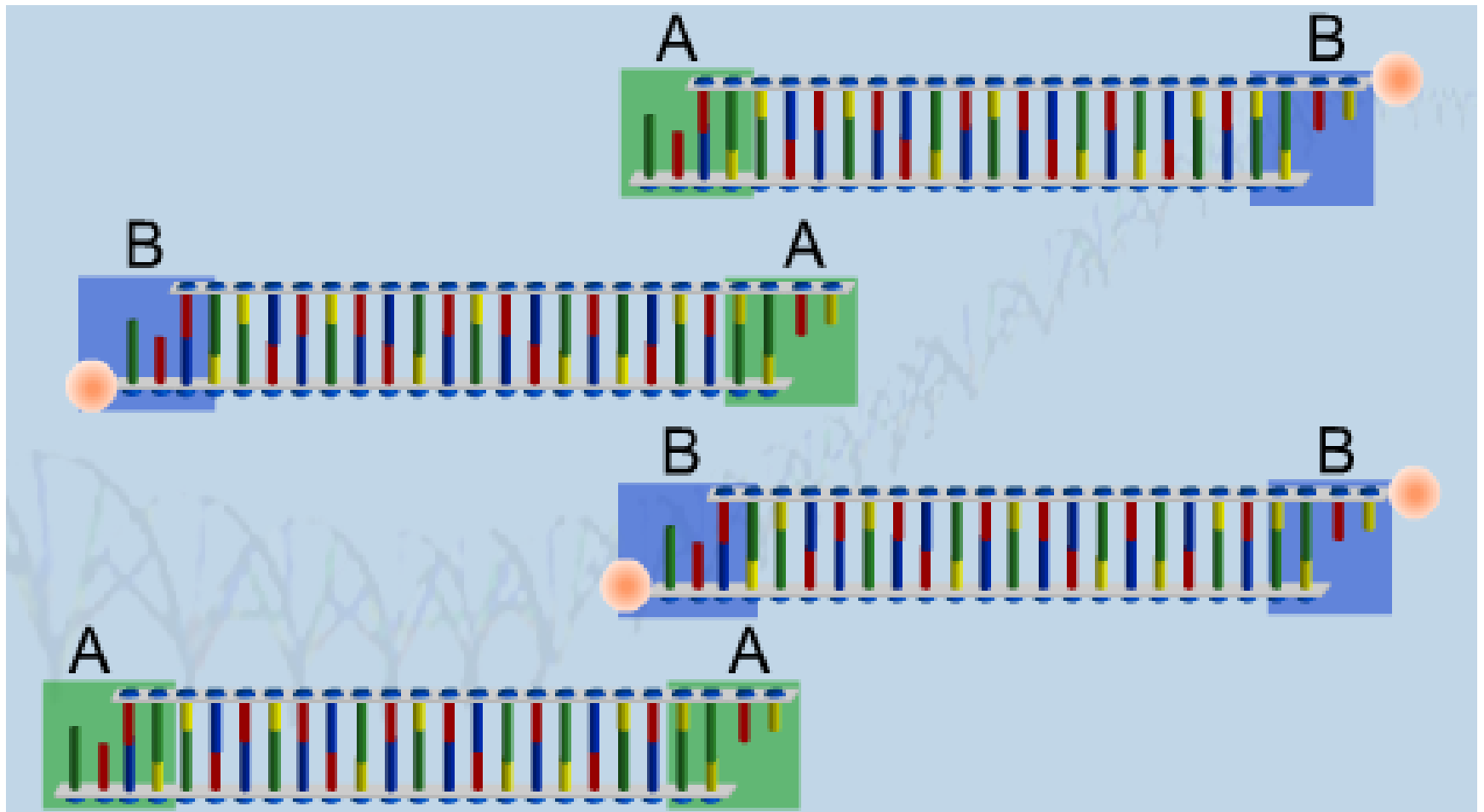
No Colony Picking



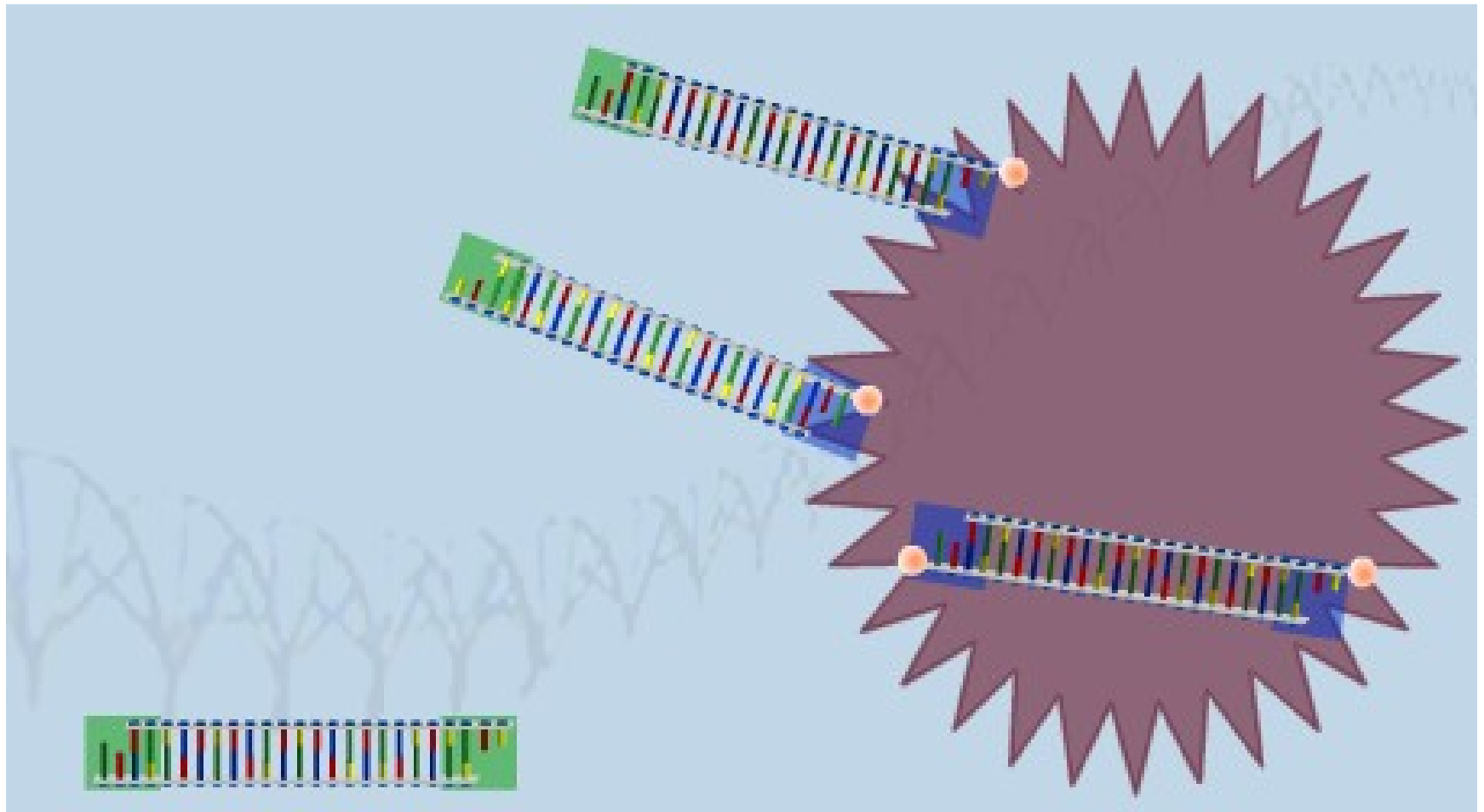
Fragmentace DNA



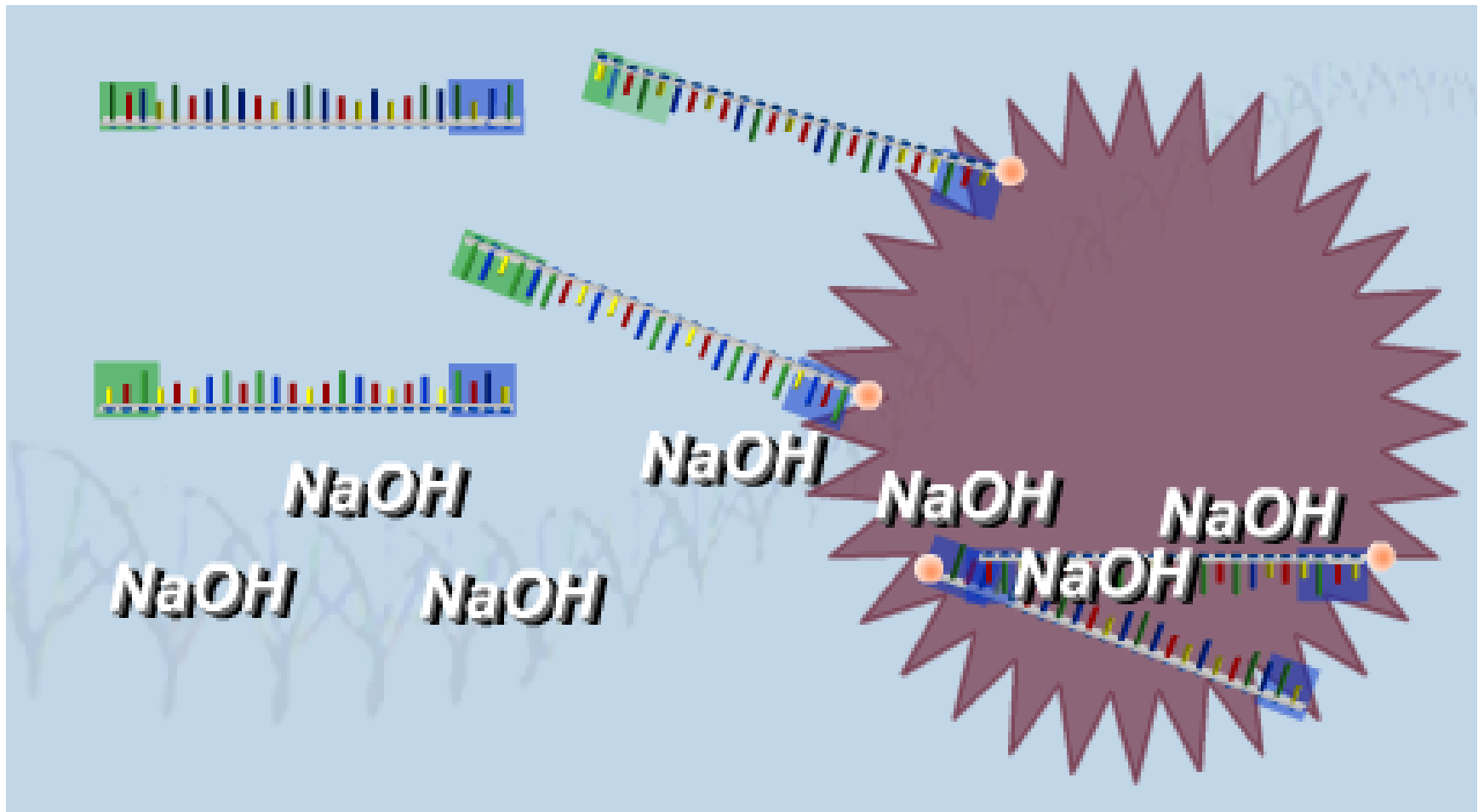
Ligace adaptoru

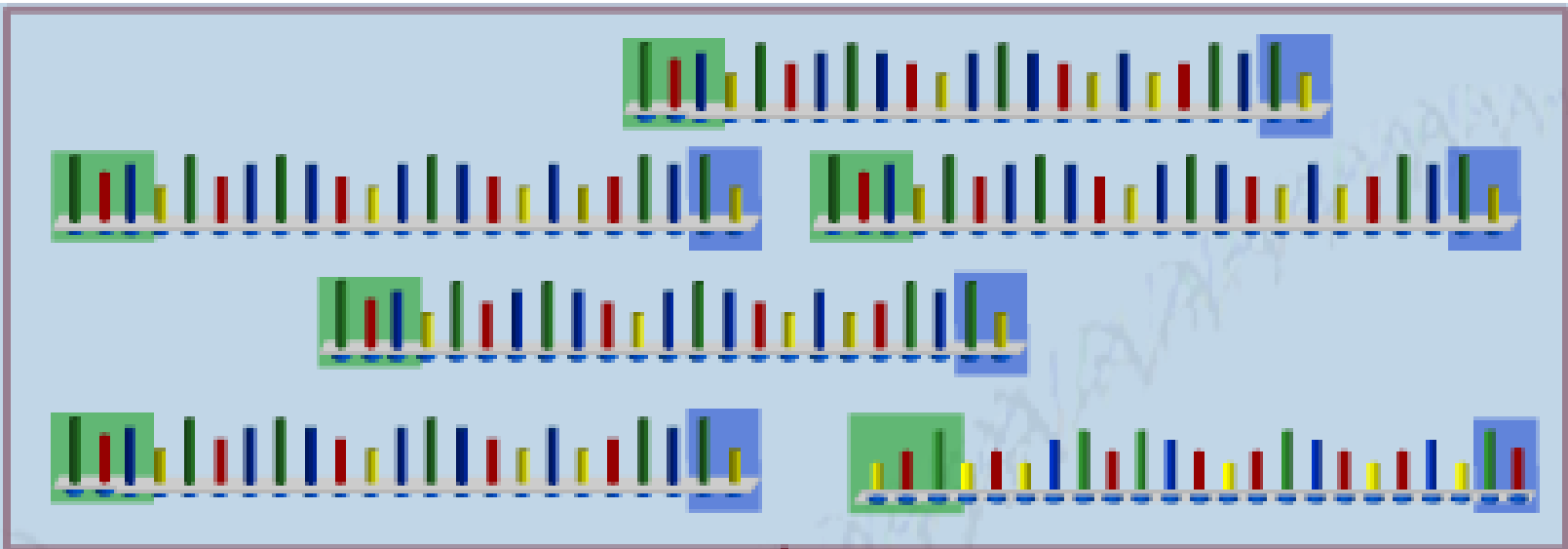


Vychytání DNA molekul



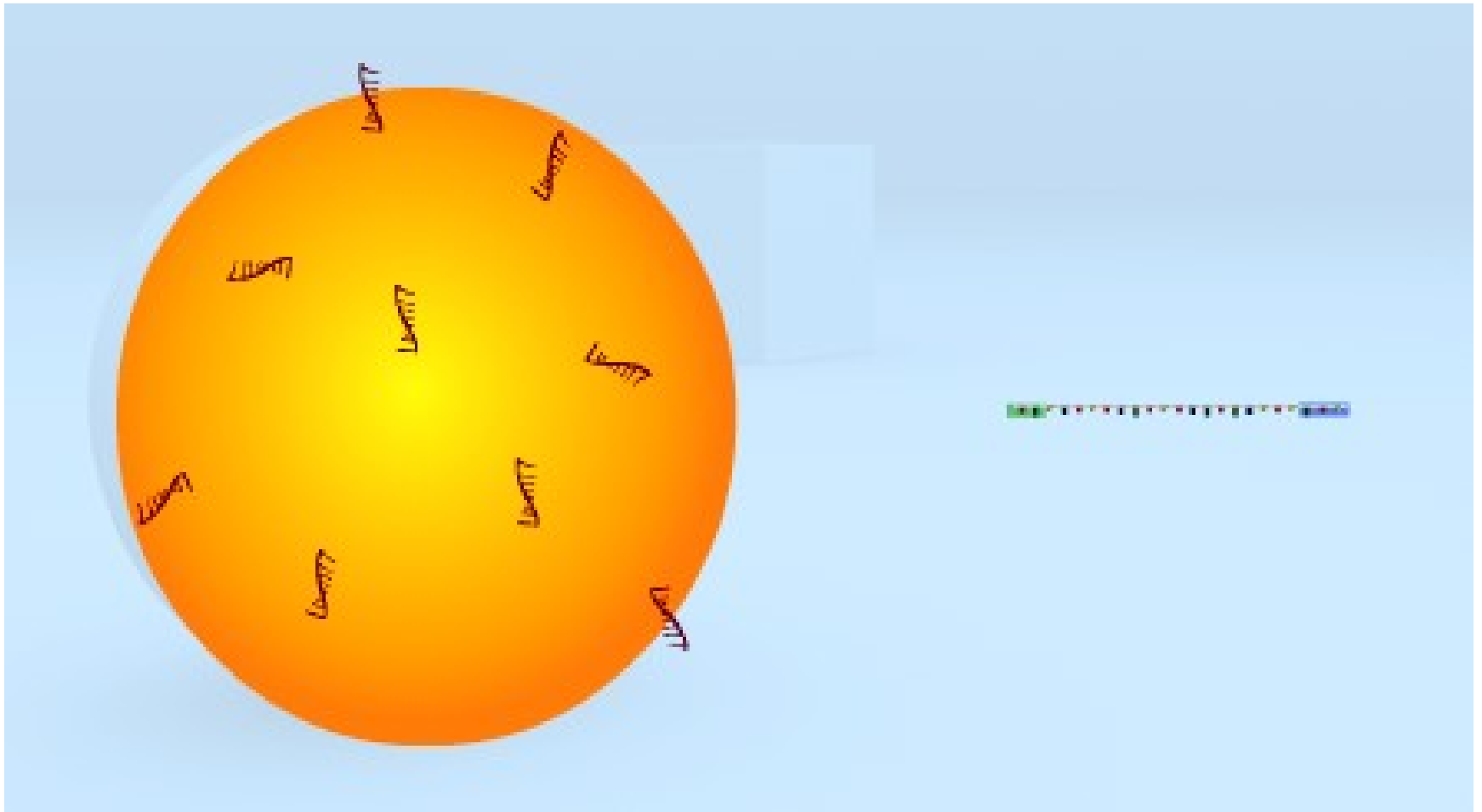
denaturace



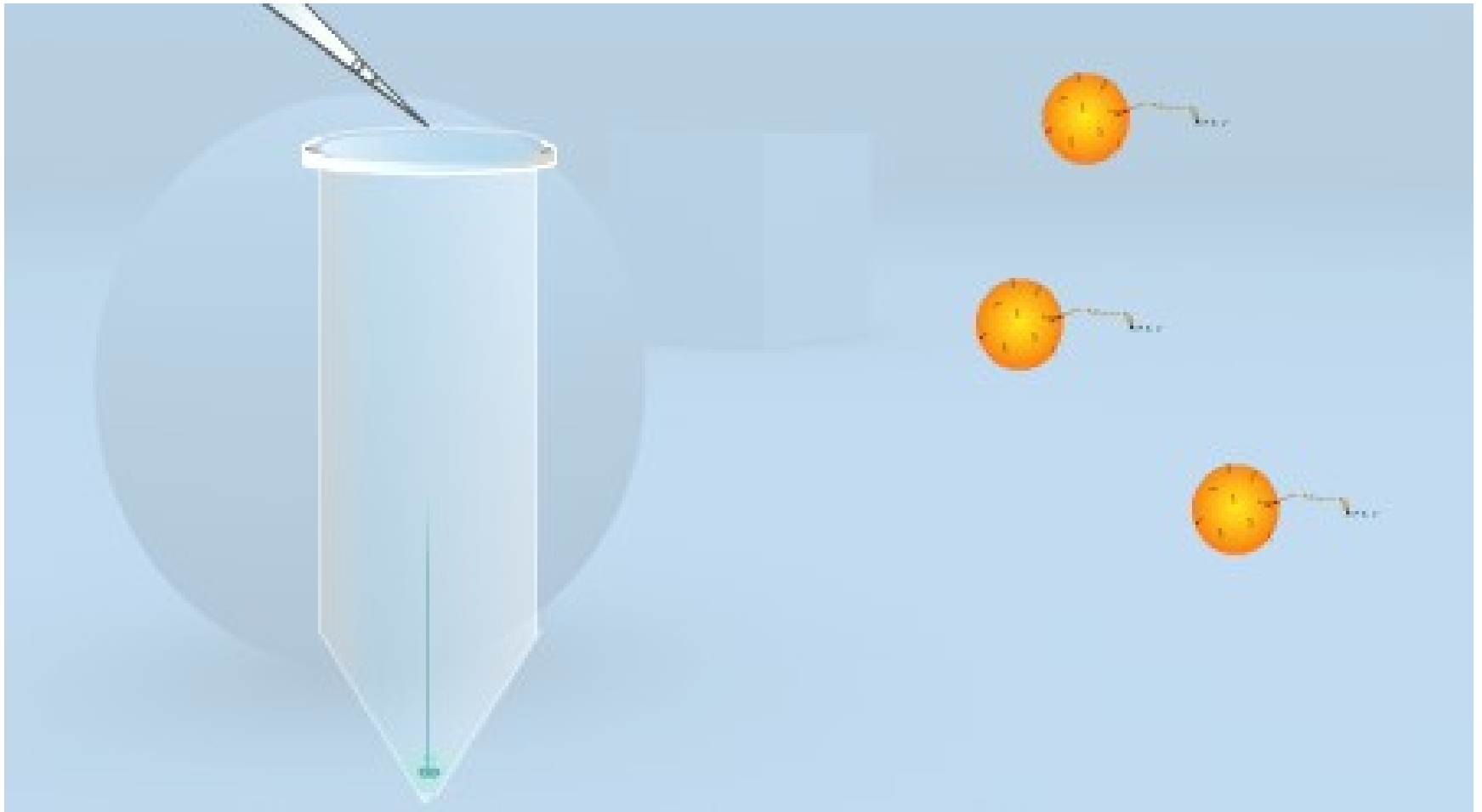


Single-stranded
template DNA
sstDNA

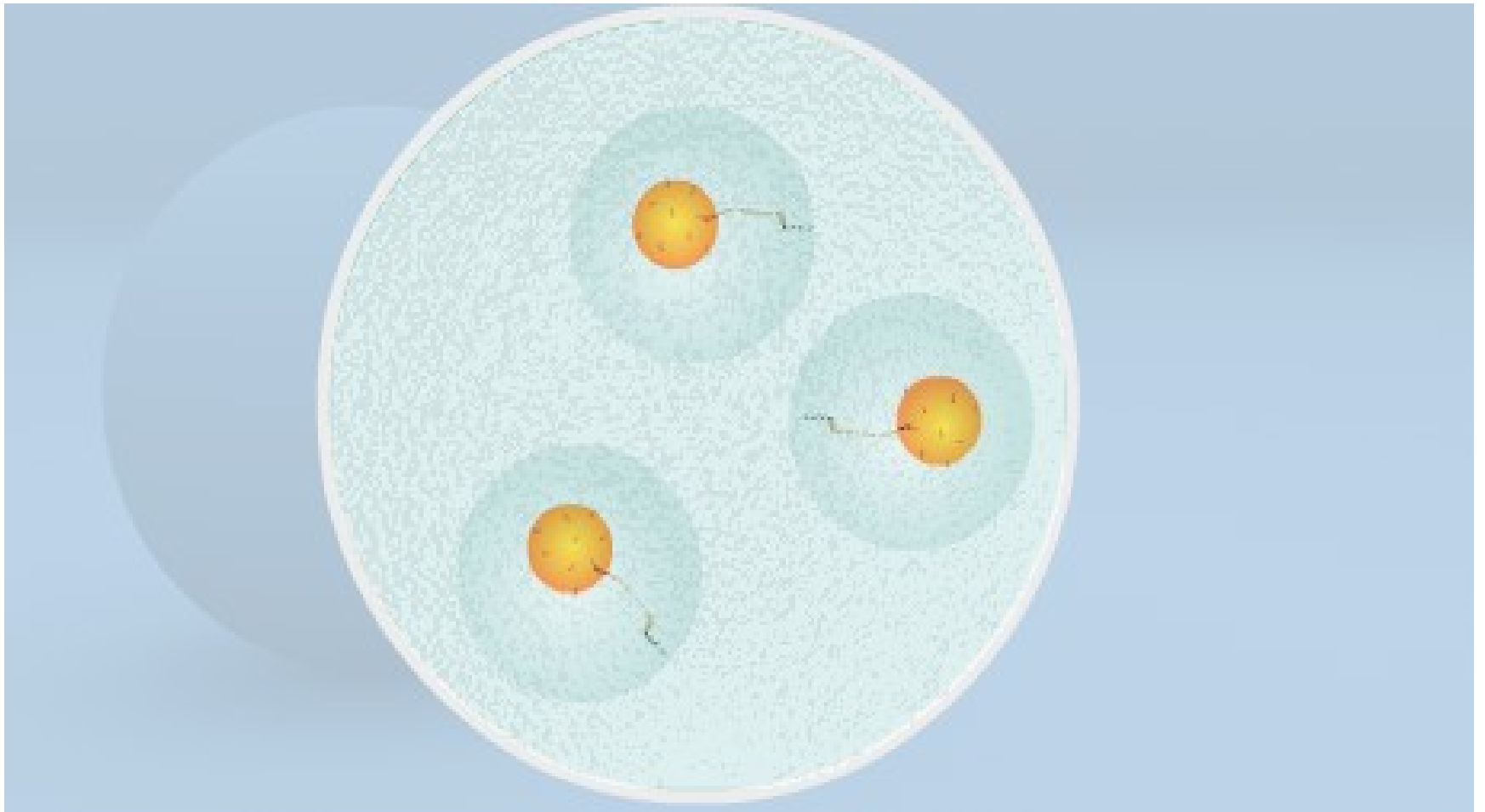
emPCR



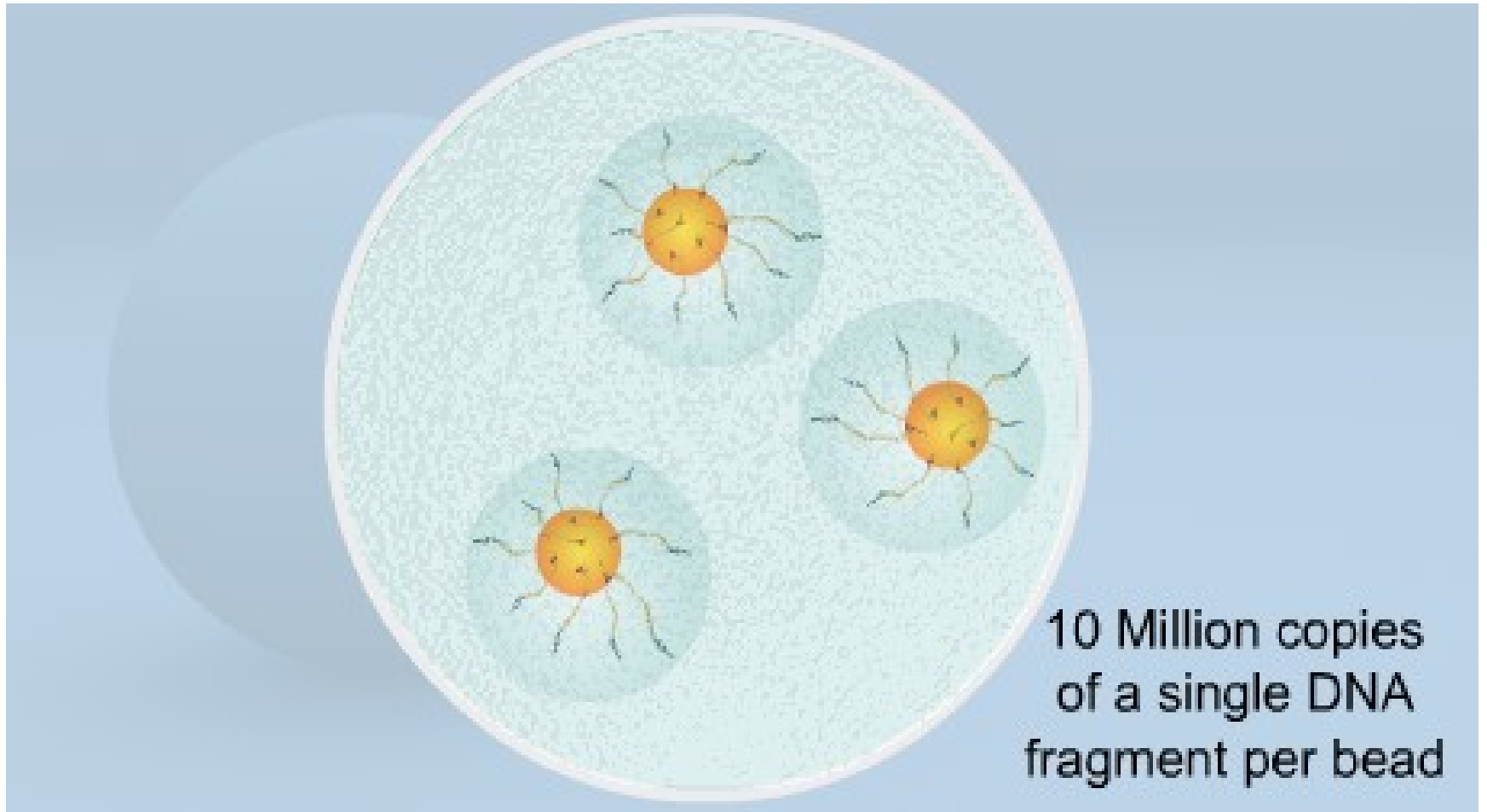
Vznik emulze (olej)



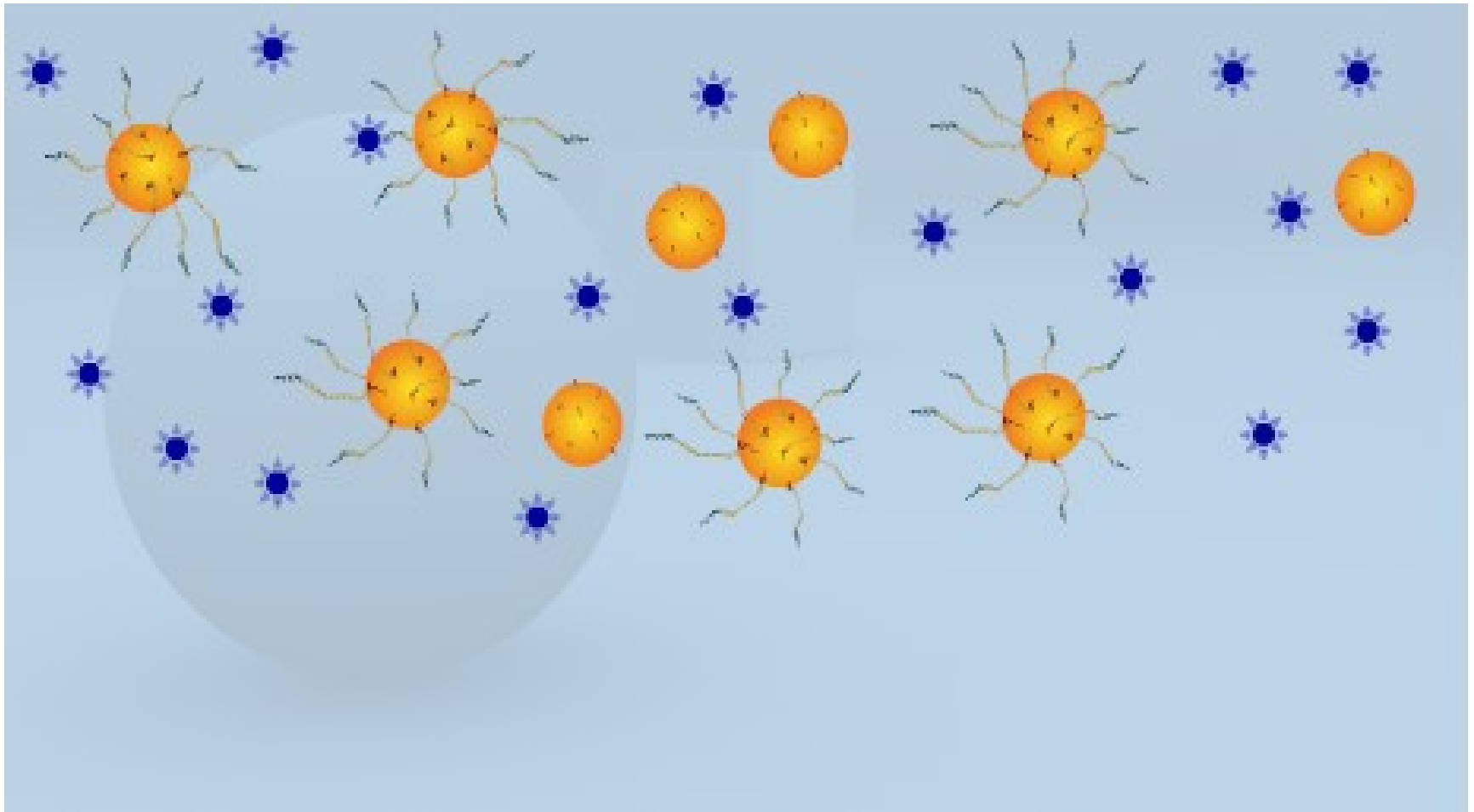
emPCR



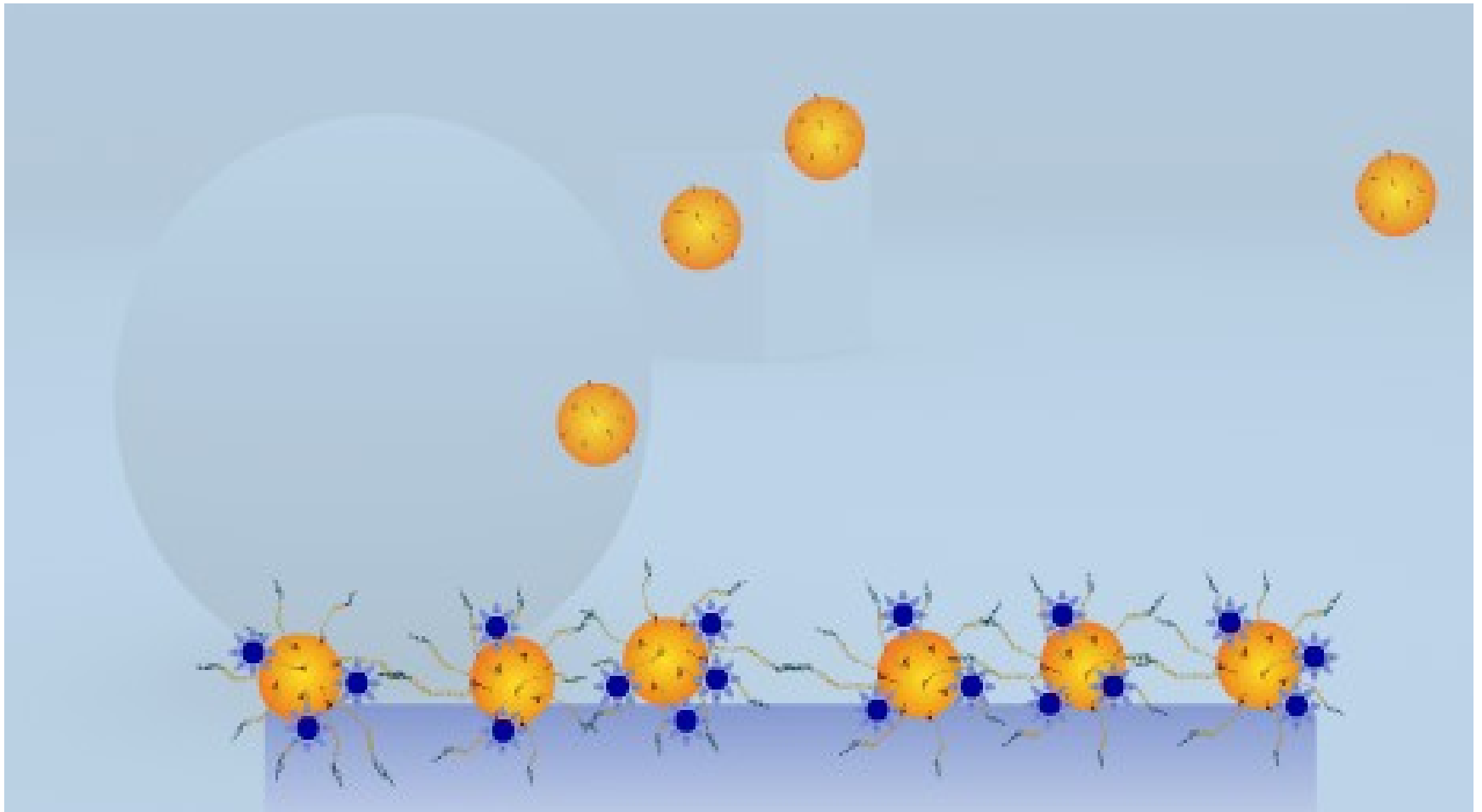
emPCR



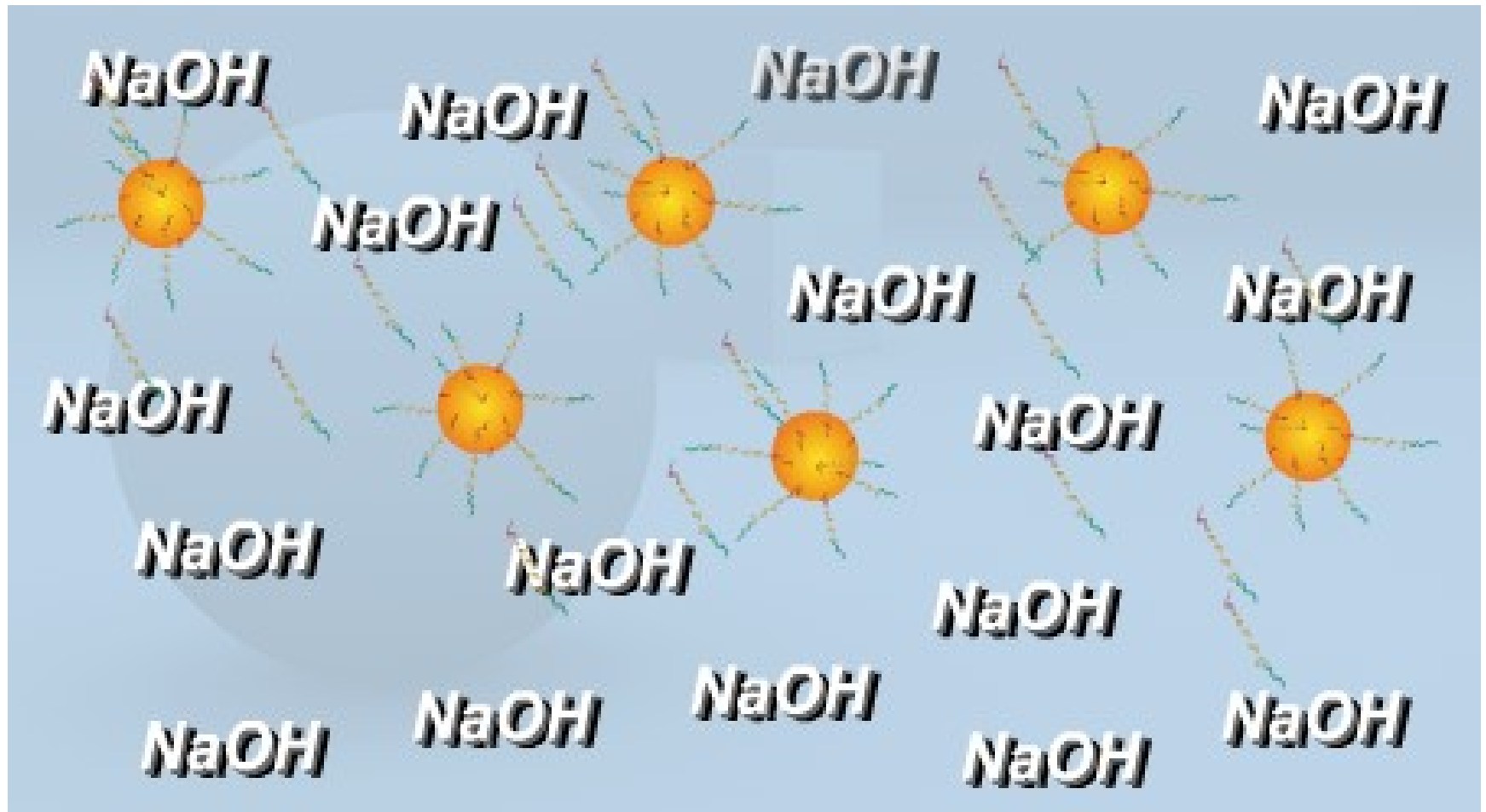
Vychytání kuliček



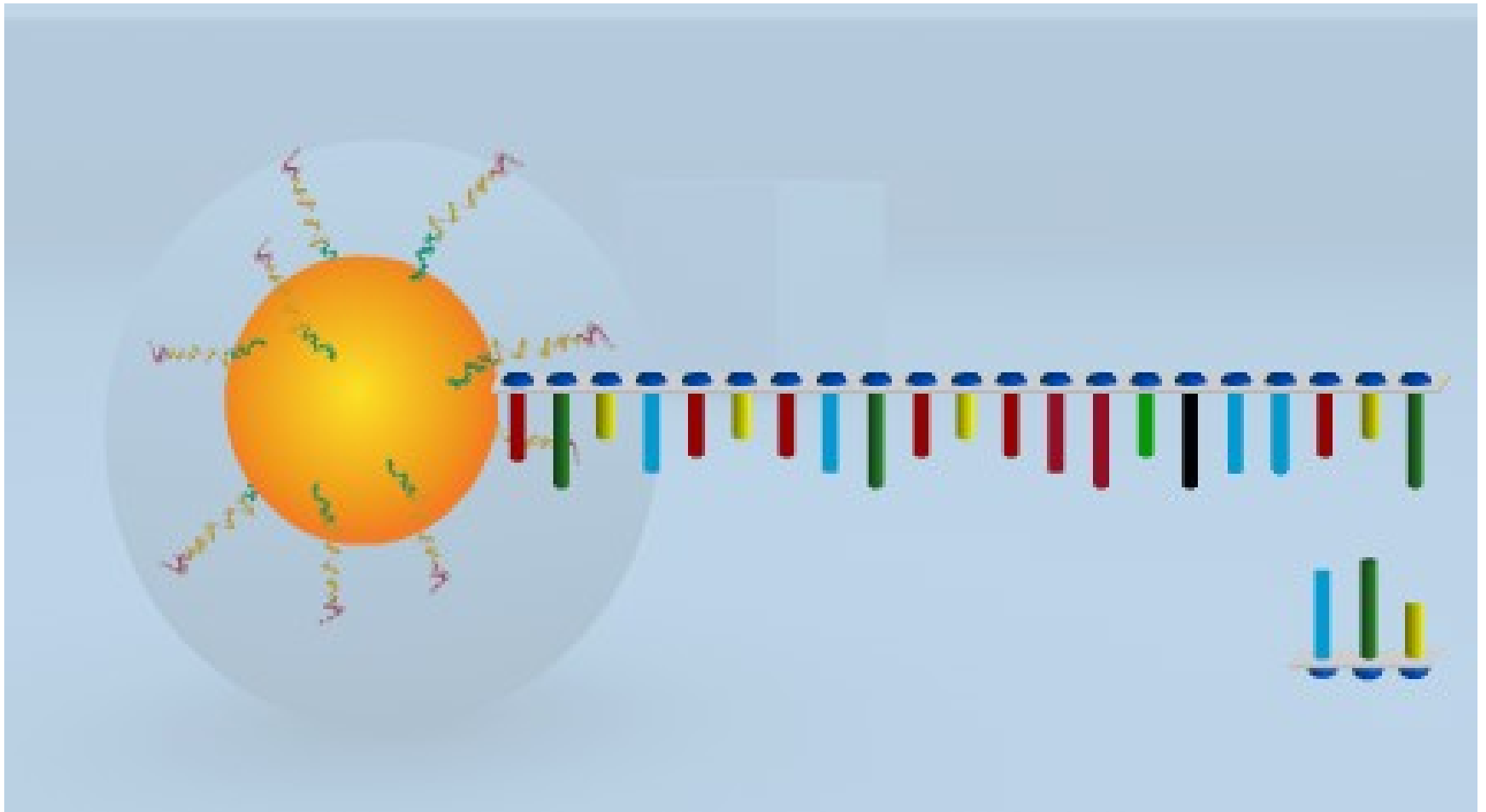
Vychytání kuliček



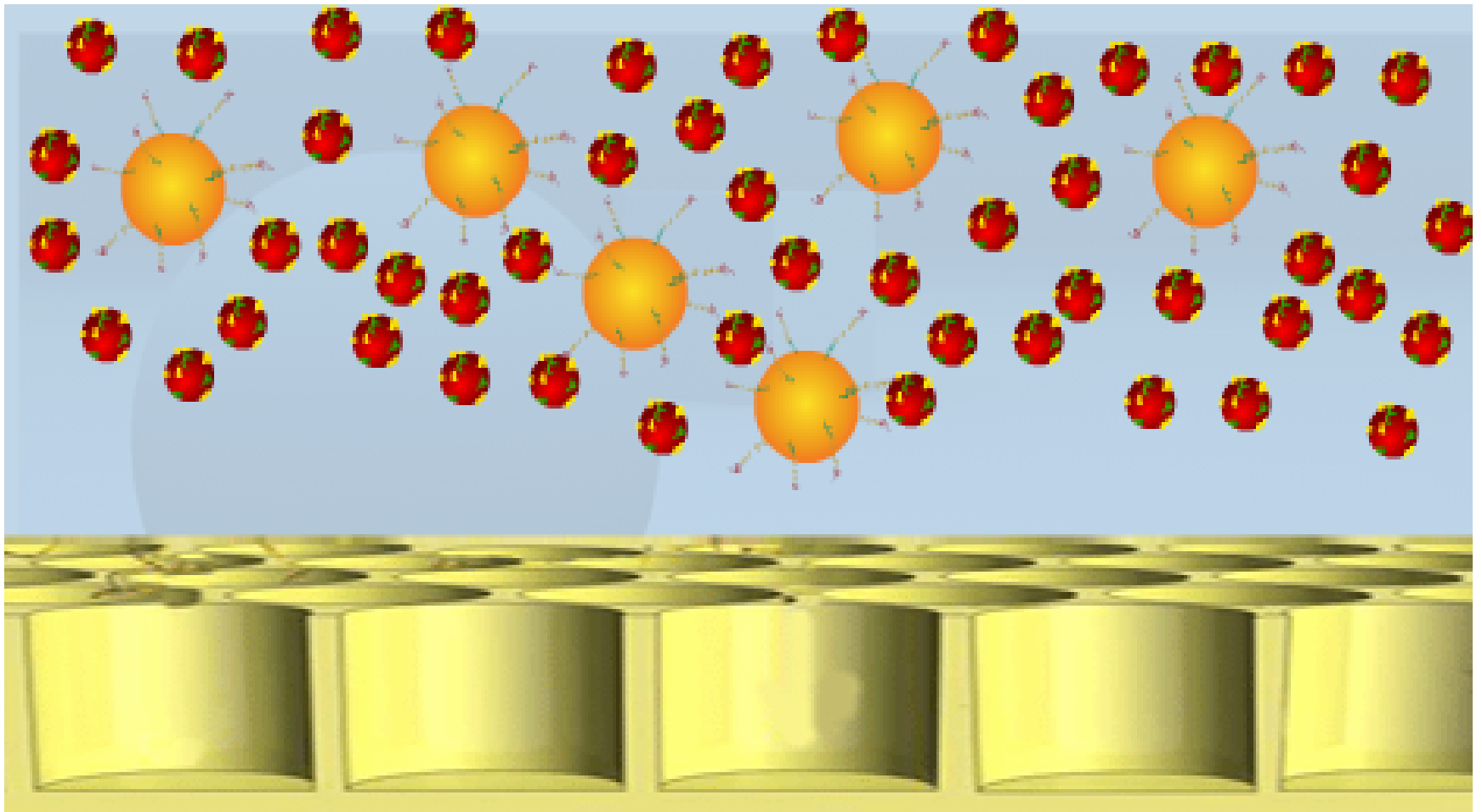
denaturace



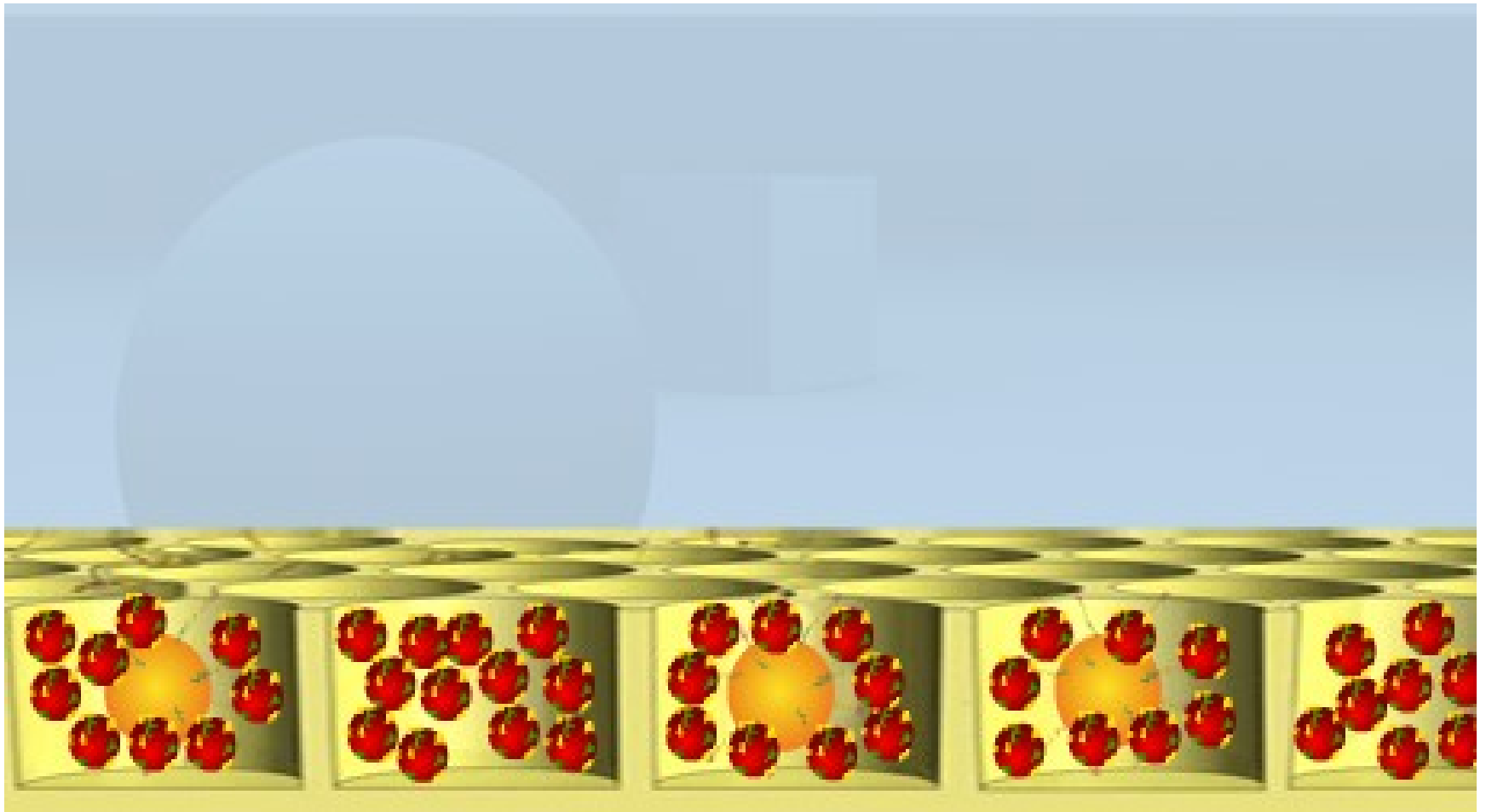
Sekvenační primer



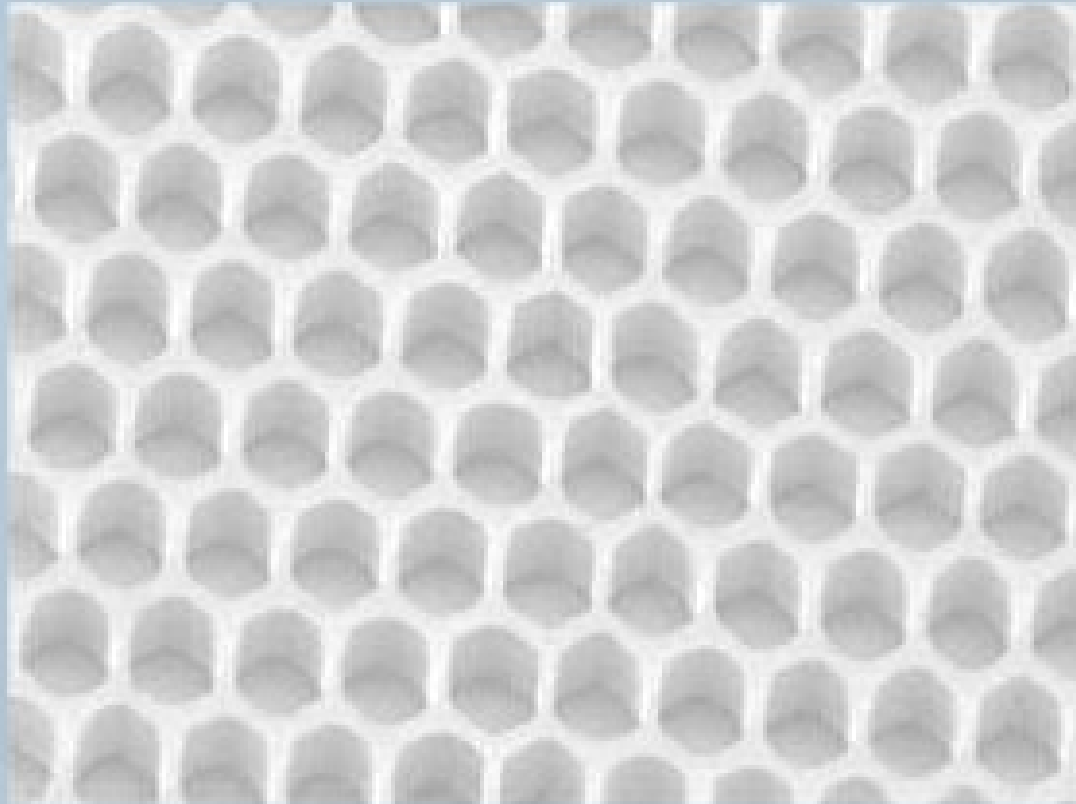
Disperze na sklíčko



Disperze na sklíčko



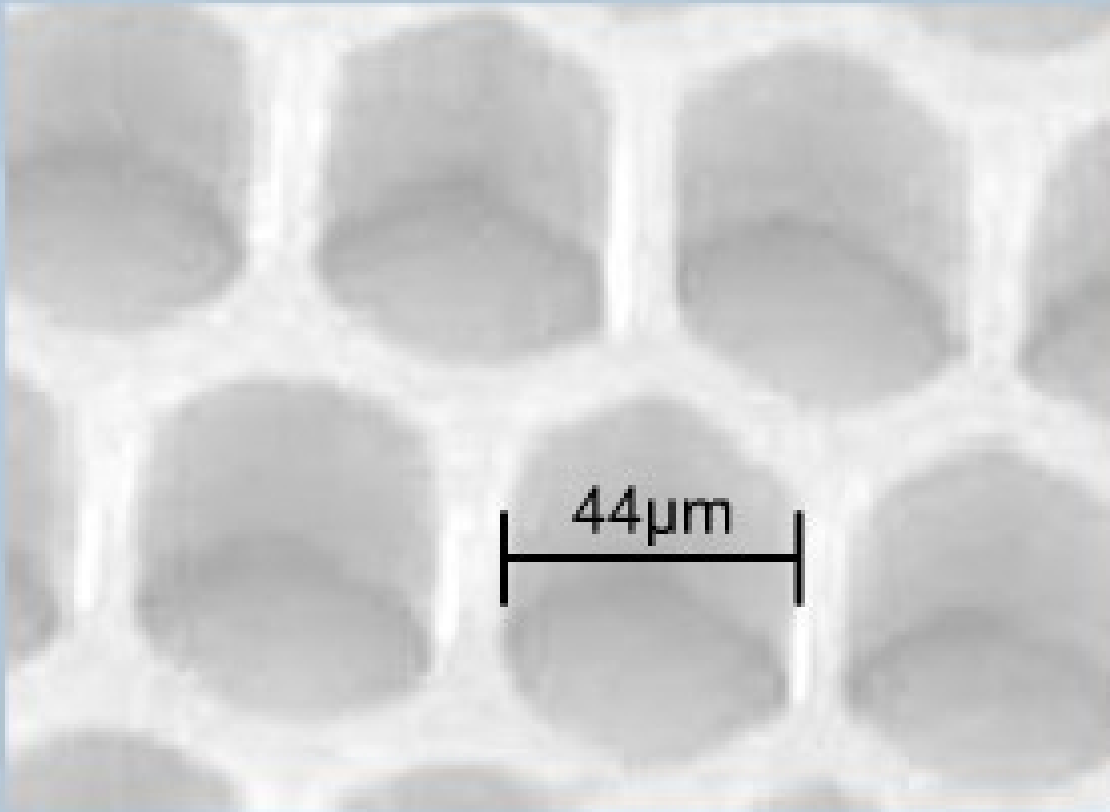
Parametry mikroreaktorů



1.6 million
wells

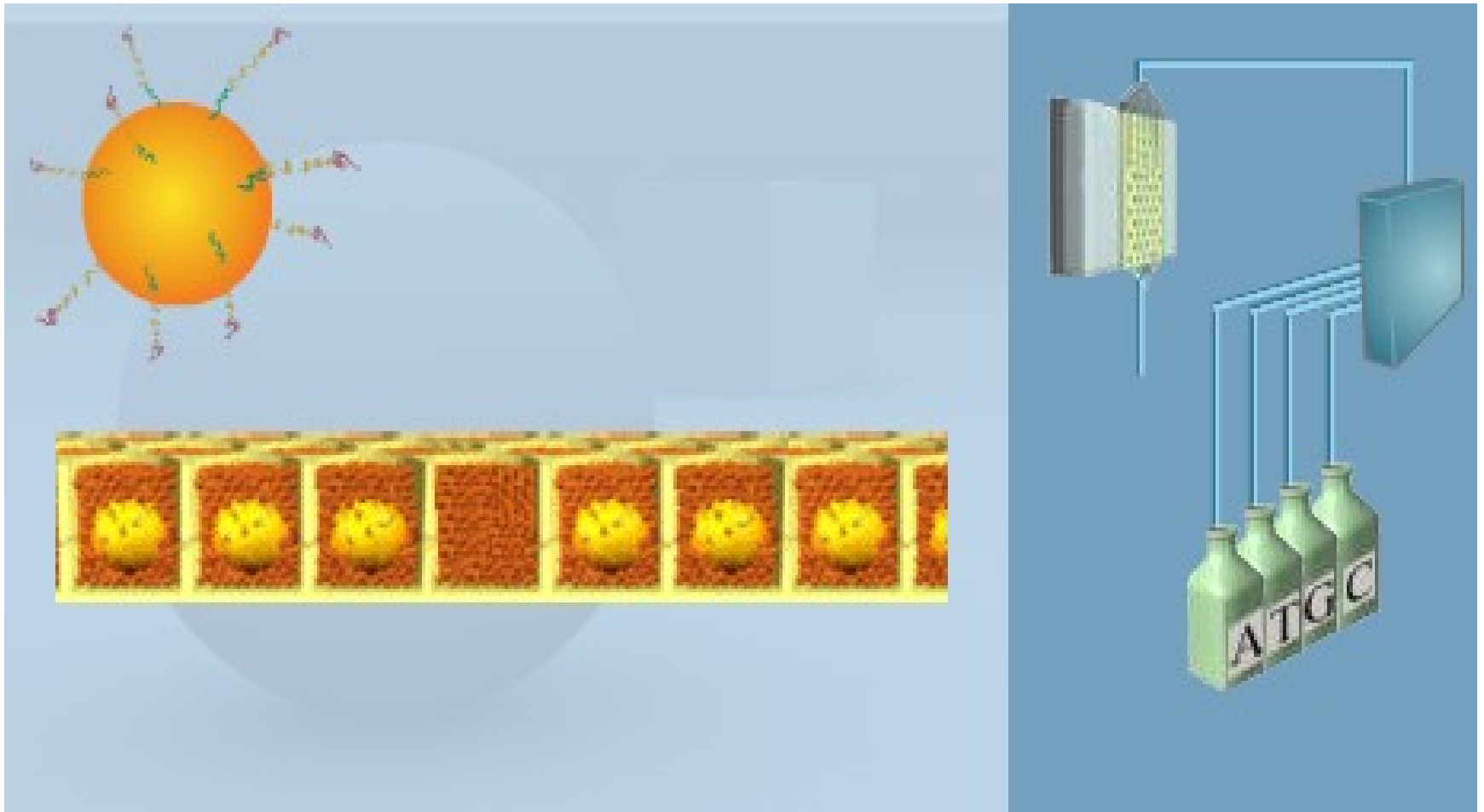
PicoTiterPlate device

Parametry mikroreaktorů

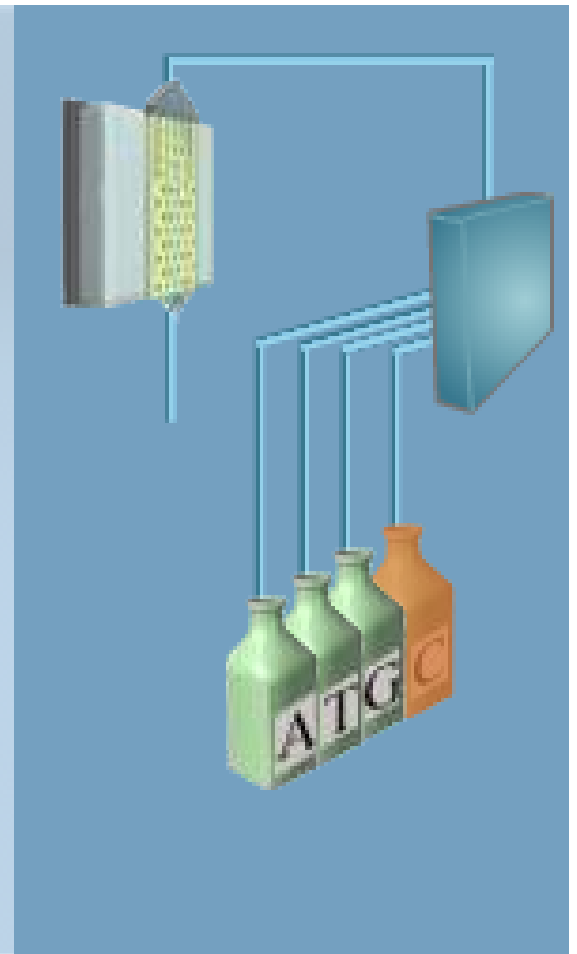
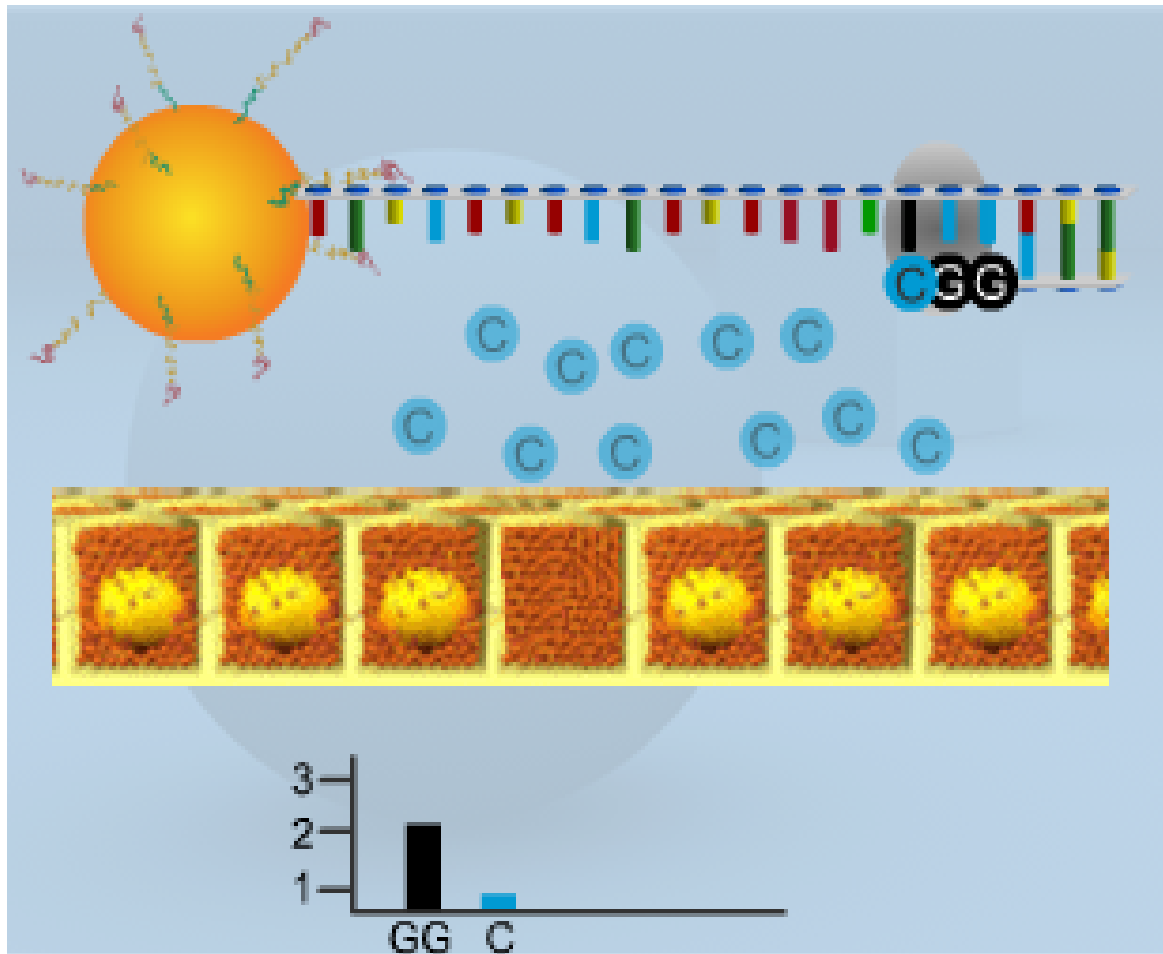


PicoTiterPlate device

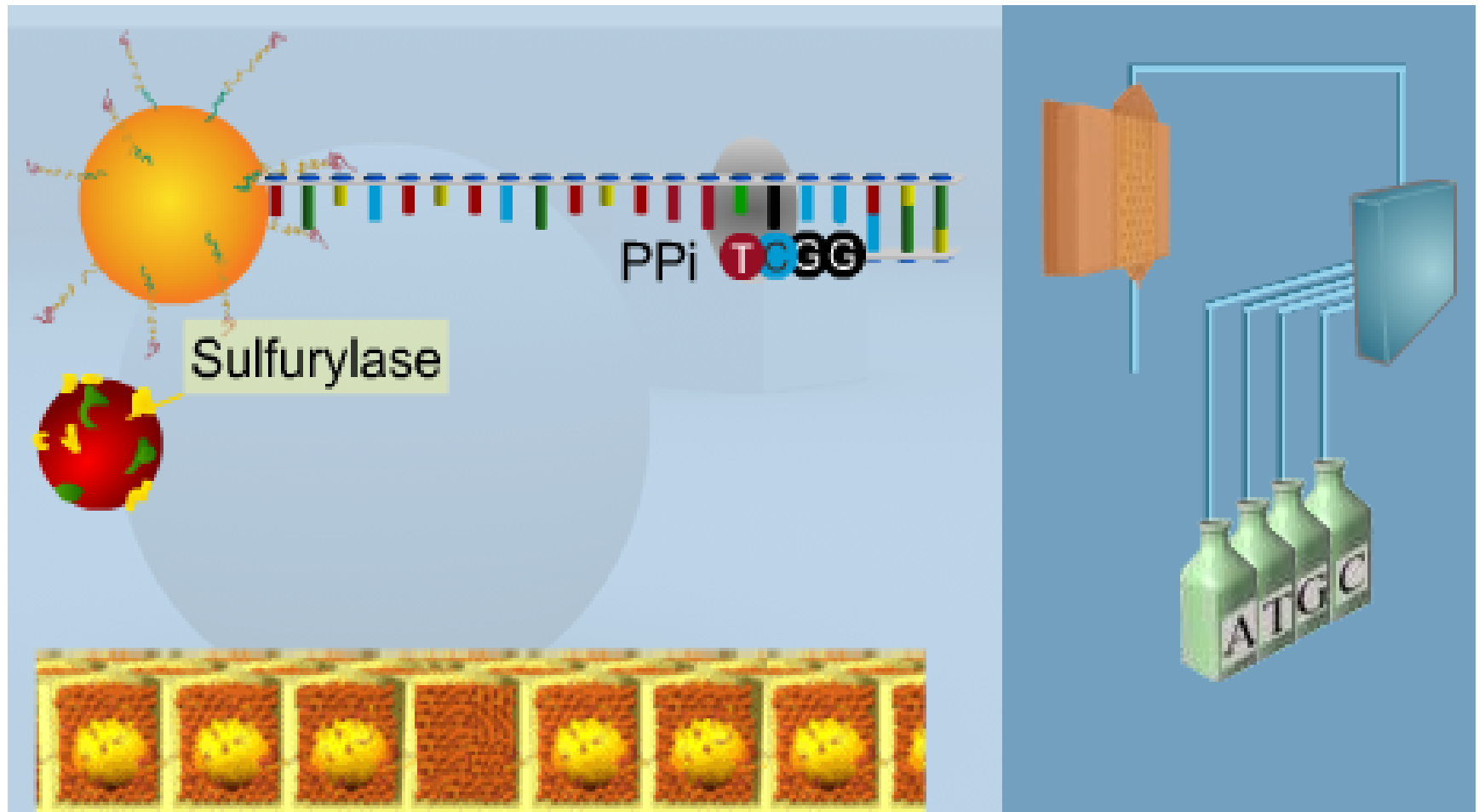
sekvenace



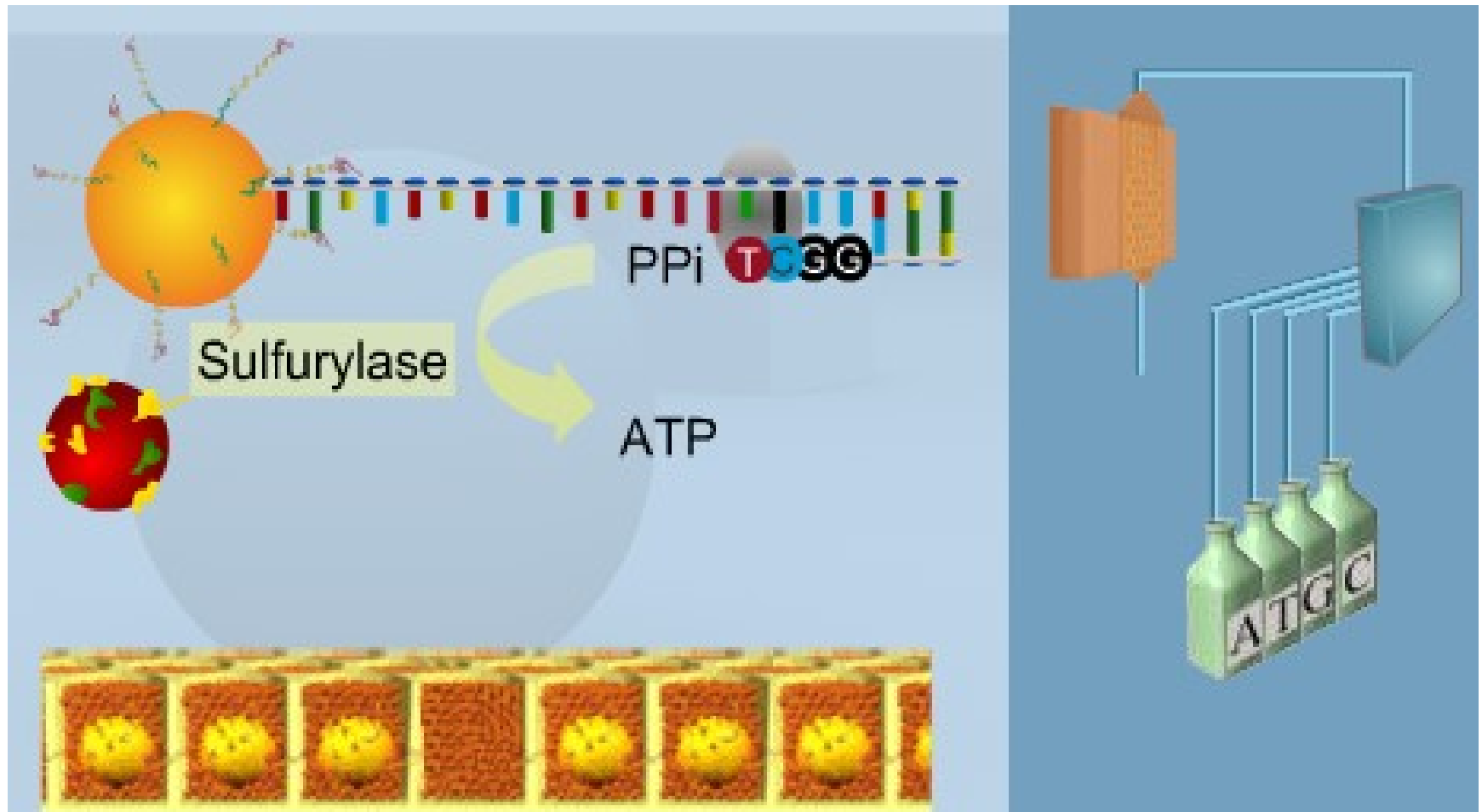
sekvenace



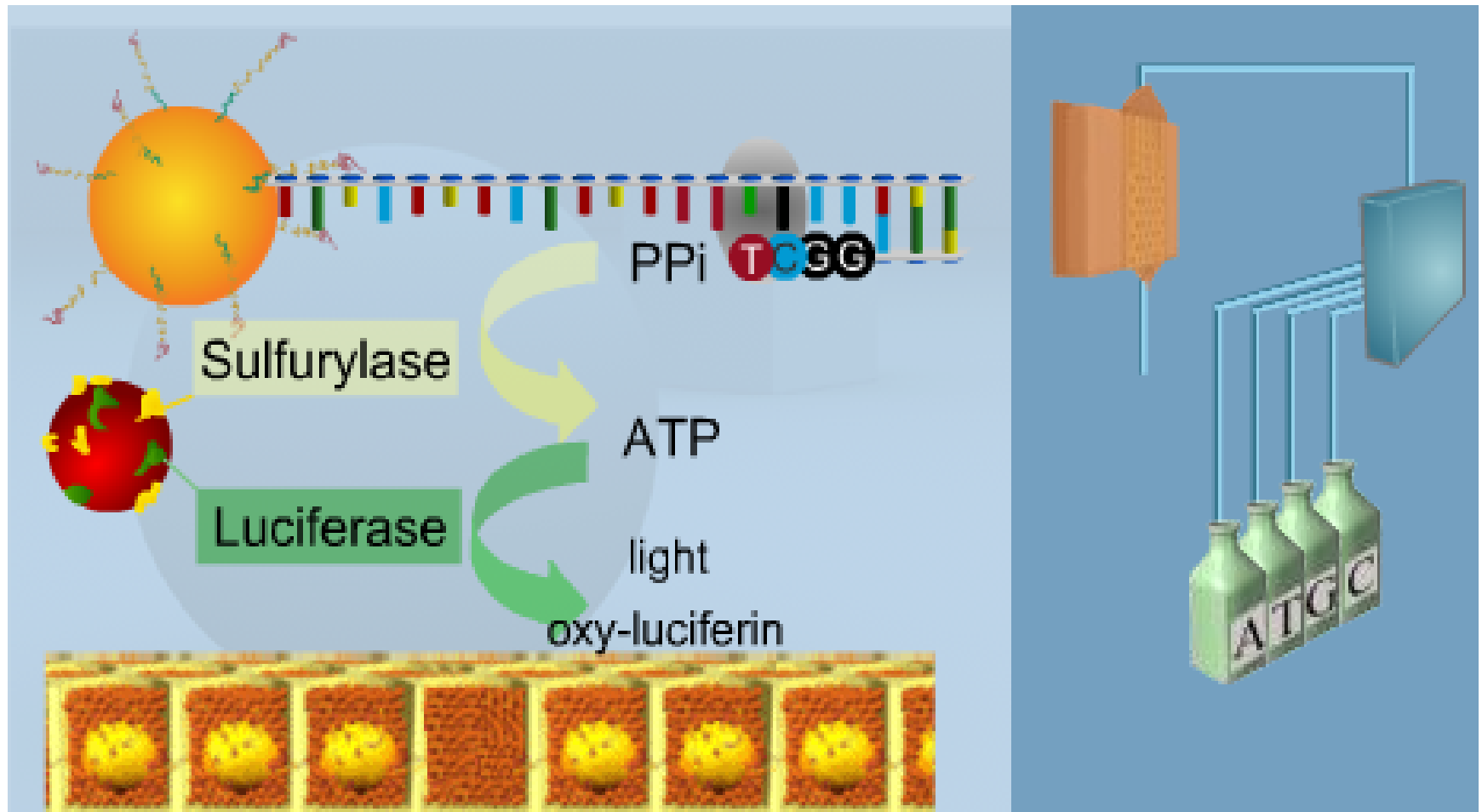
sekvenace



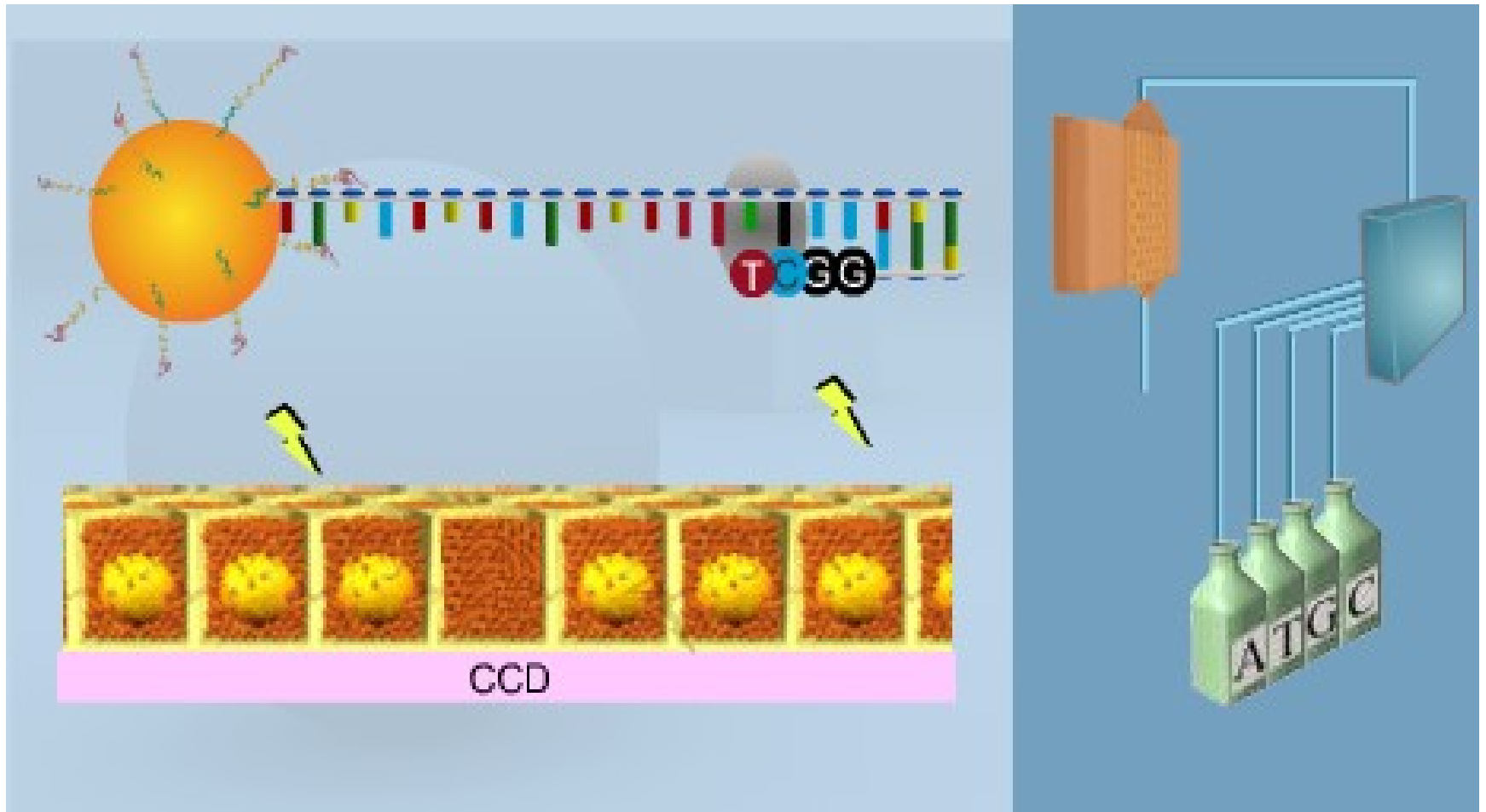
sekvenace



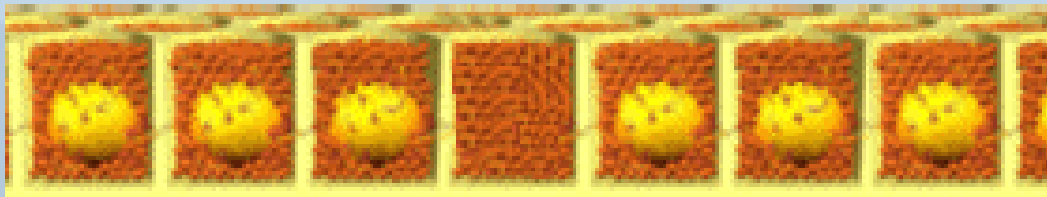
sekvenace



sekvenace



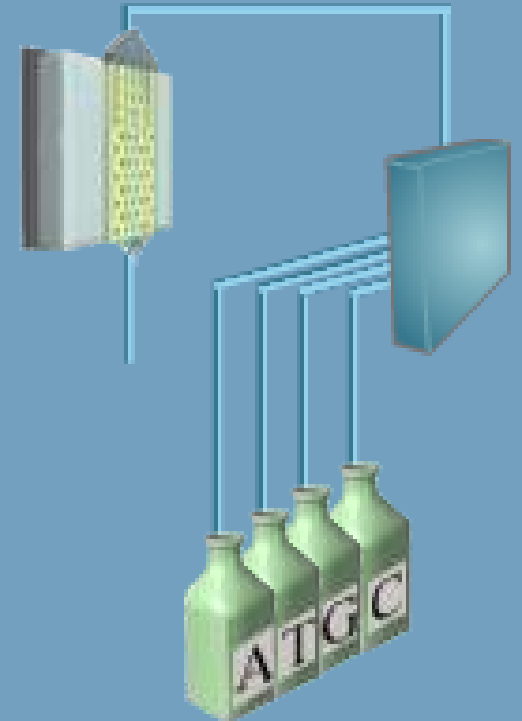
sekvenace



Massive parallelization of sequencing reactions

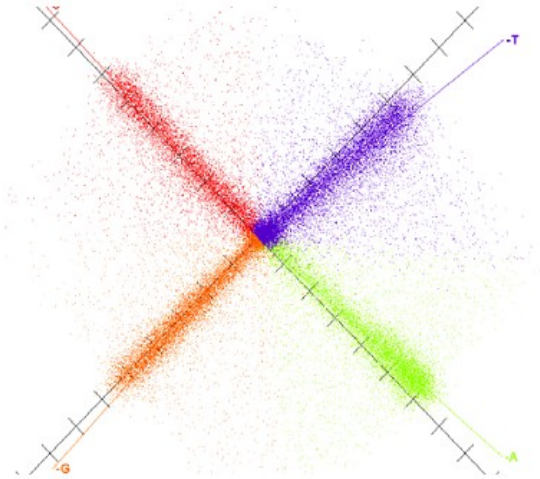
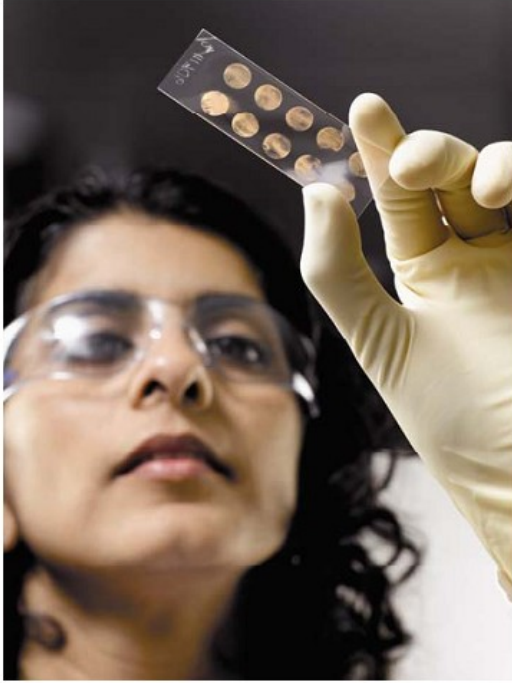
100 bases read length
X
200 000

20 Million Bases



SOLID (Sequencing by Oligonucleotide Ligation and Detection)

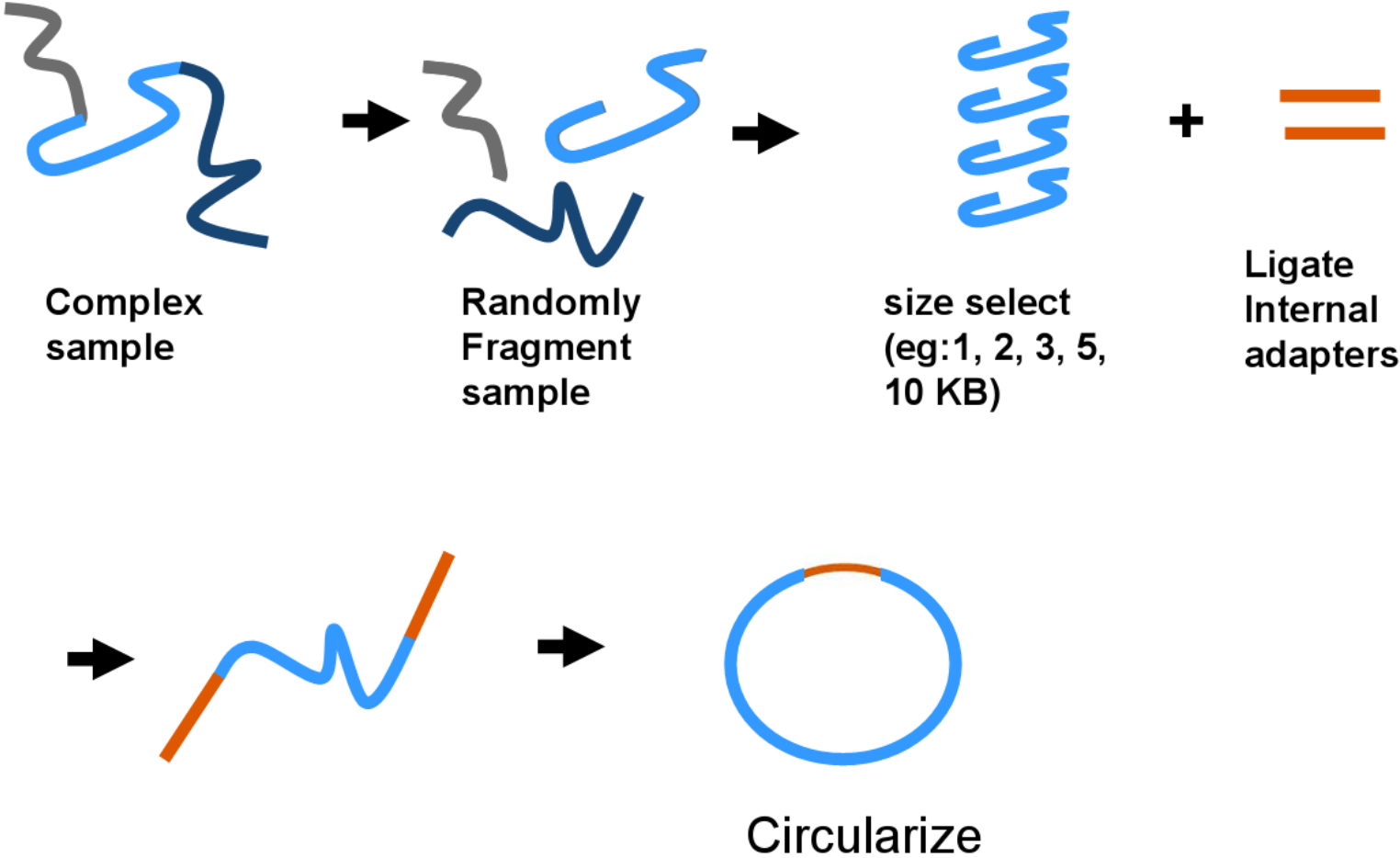
2-base encoding sequencing (2007)



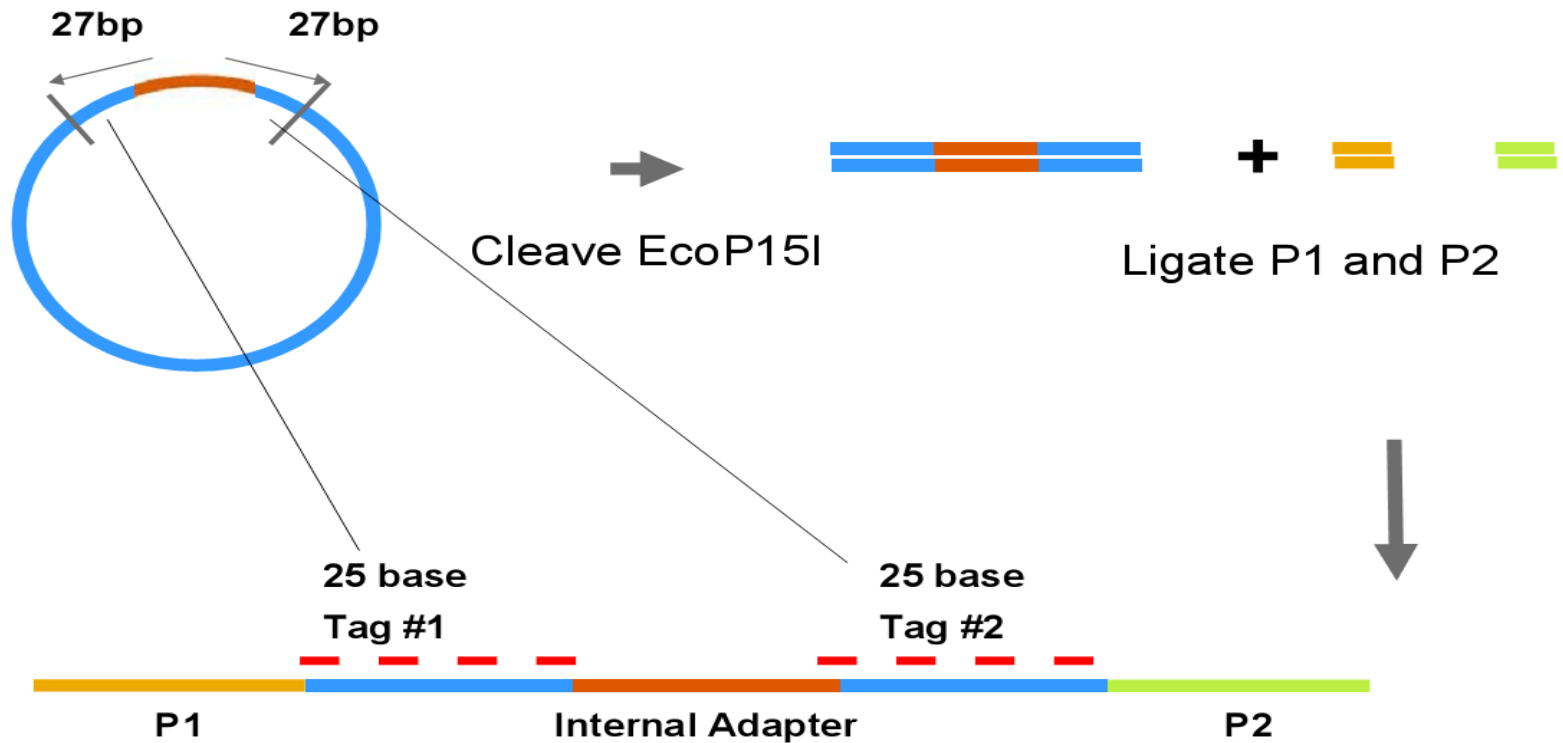
SOLiD™ System

Sequencing by **O**ligonucleotide **L**igation and **D**etection

Creating a mate-paired library (i)



Creating a mate-paired library (ii)



EcoP15I



[Nomenclature Update](#)

Catalog #	Size	Concentration	Price	Qty	
R0646S	25 units	2,000 units/ml	\$61.00	<input type="text" value="1"/>	ADD TO CART
R0646L	125 units	2,000 units/ml	\$244.00	<input type="text" value="1"/>	ADD TO CART

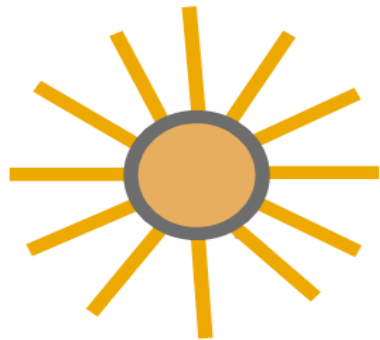
Prices are in US dollars and valid only for US orders.

Download: [MSDS PDF](#)

Recognition Site:

5'... CAGCAG(N)₂₅... 3'
3'... GTCGTC(N)₂₇... 5'

Emulsion PCR



P1-coupled beads

+



Templates



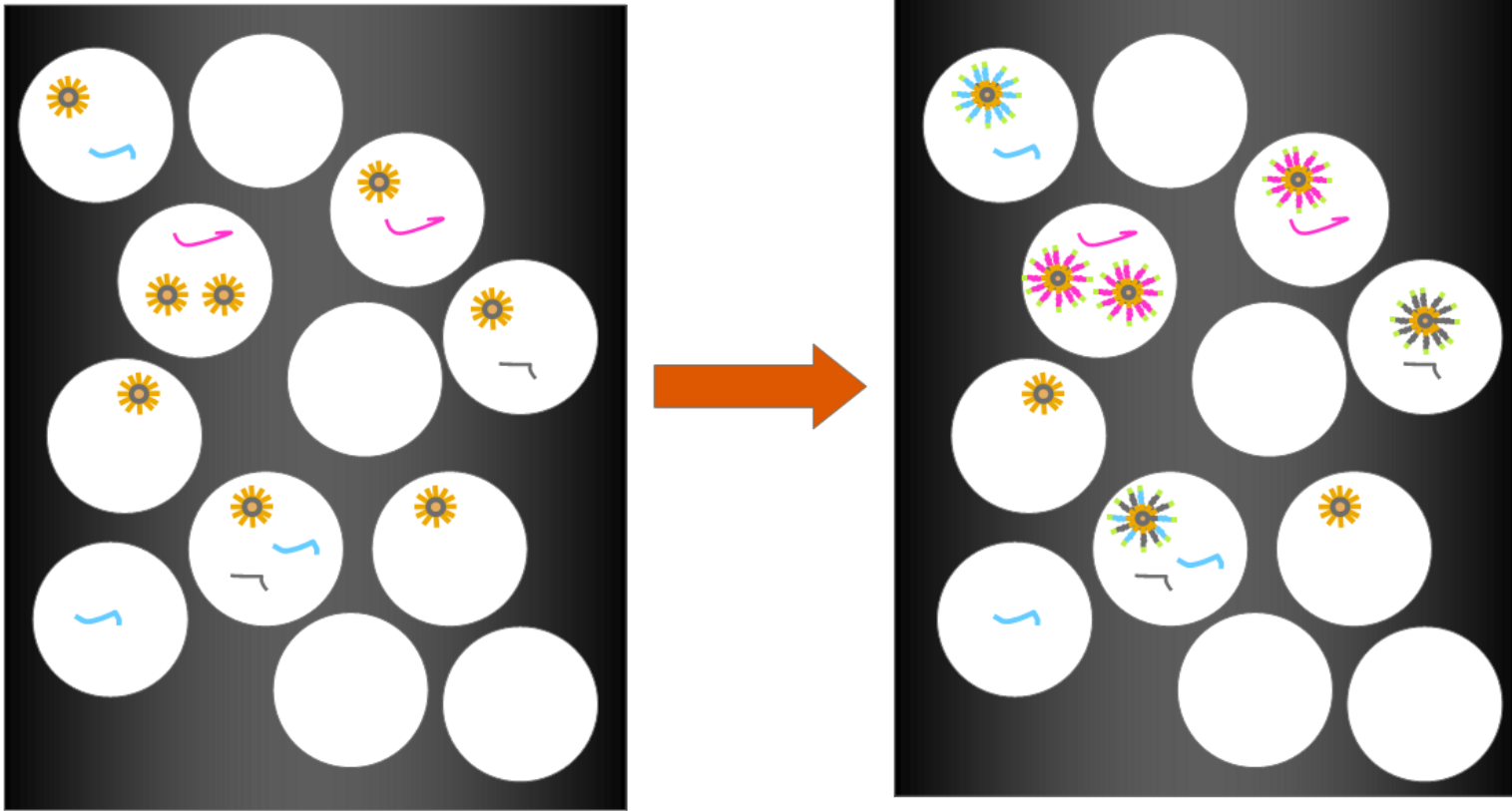
Primers P1 << P2



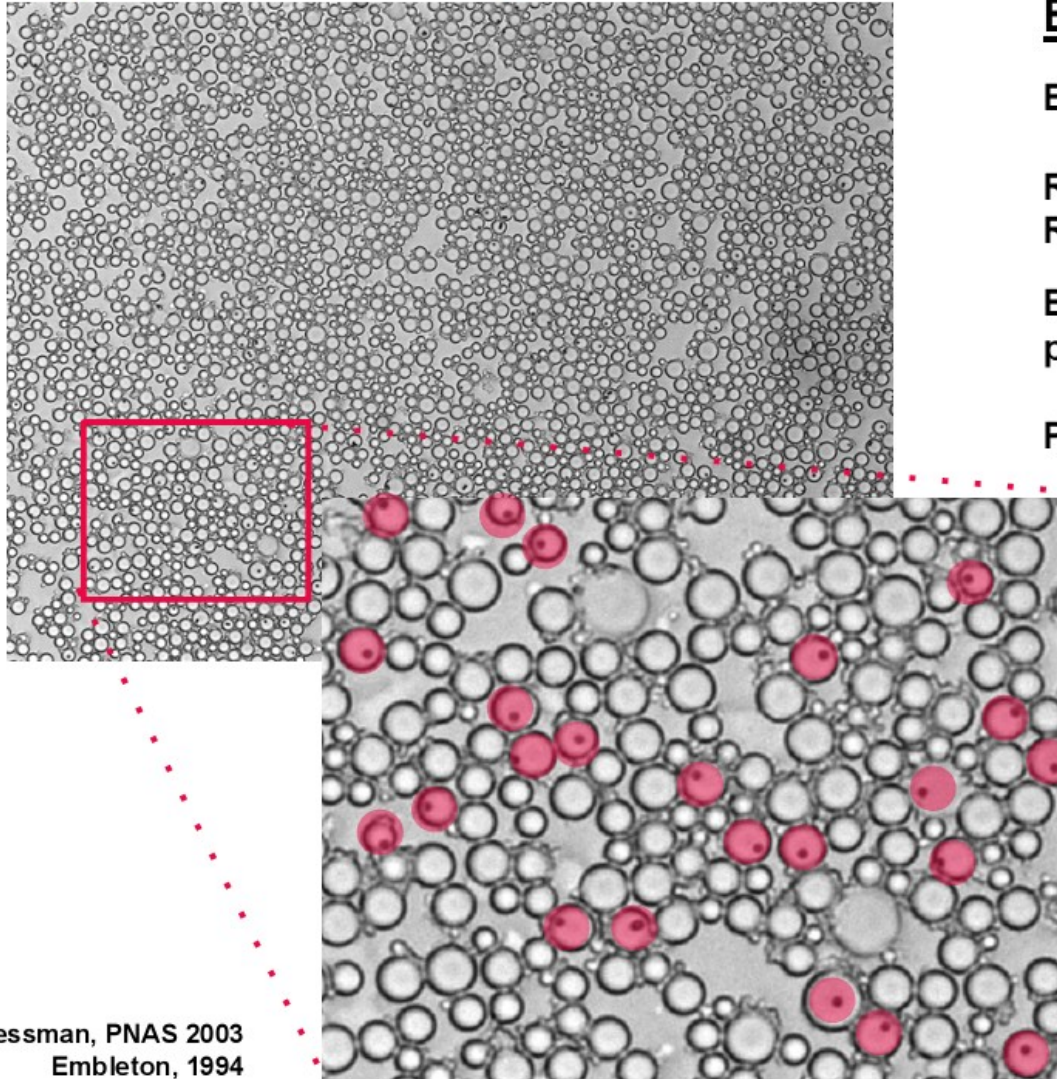
Enzyme

Emulsion PCR

Mix PCR aqueous phase into a water-in-oil (w/o) emulsion and carry out emulsion PCR



SOLiD system: Clonal Amplification



Emulsion Metrics

Bead size: 1 μm

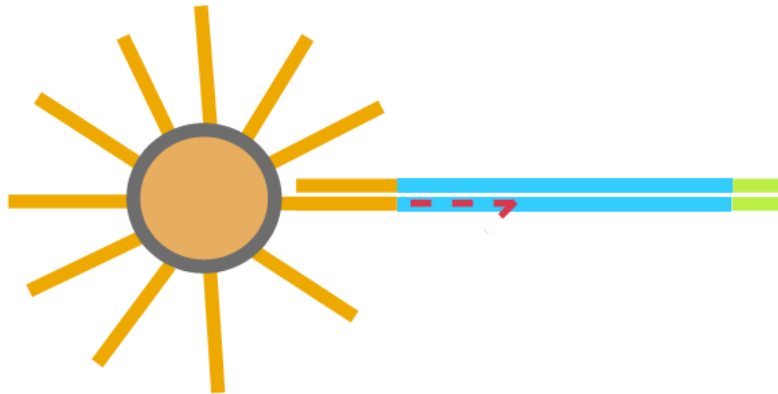
Reactor size: 4 μm

Reactor volume: 34 fL

Beads / emulsion
plate (96-well): $2-4 \times 10^9$

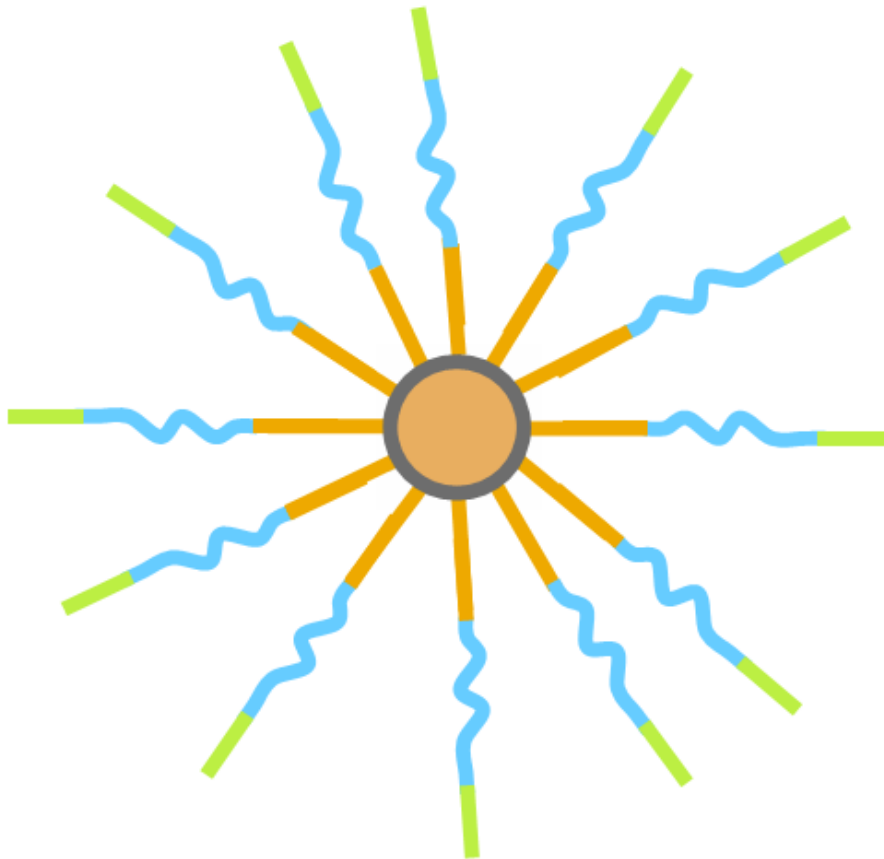
Post Enrichment: $\sim 500\text{M}$ / plate

emulsion PCR Individual Bead

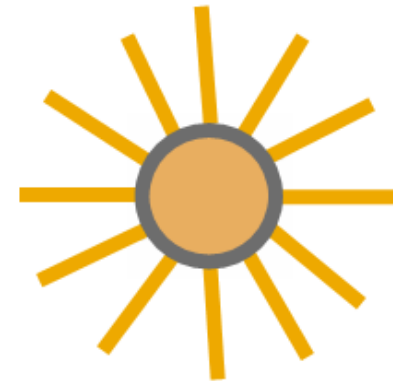


- 1) Template Anneals to P1
- 2) Polymerase extends from P1
- 3) Complementary sequence is extended off bead surface
- 4) Template disassociates

Results of emulsion PCR, after breaking emulsion

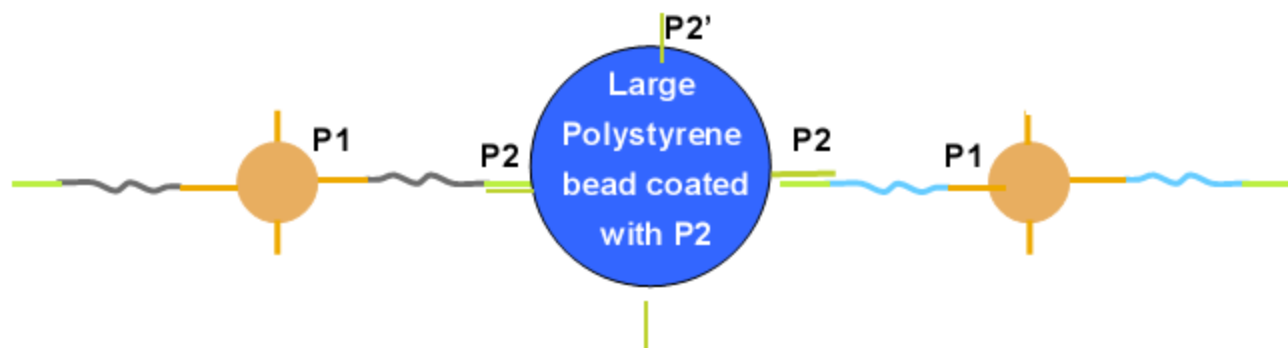


Bead contains ~20K amplified products from original single strand molecule

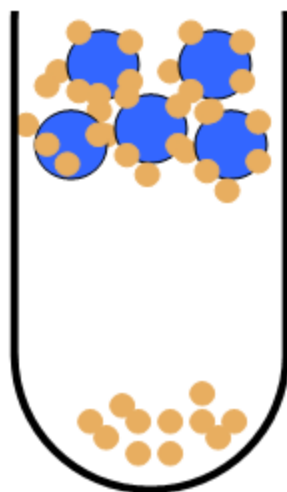


Beads with no product

Enrichment

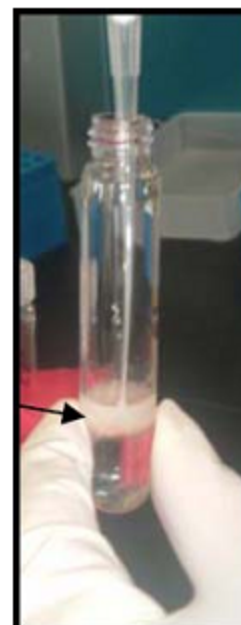


**Centrifuge in
glycerol gradient**

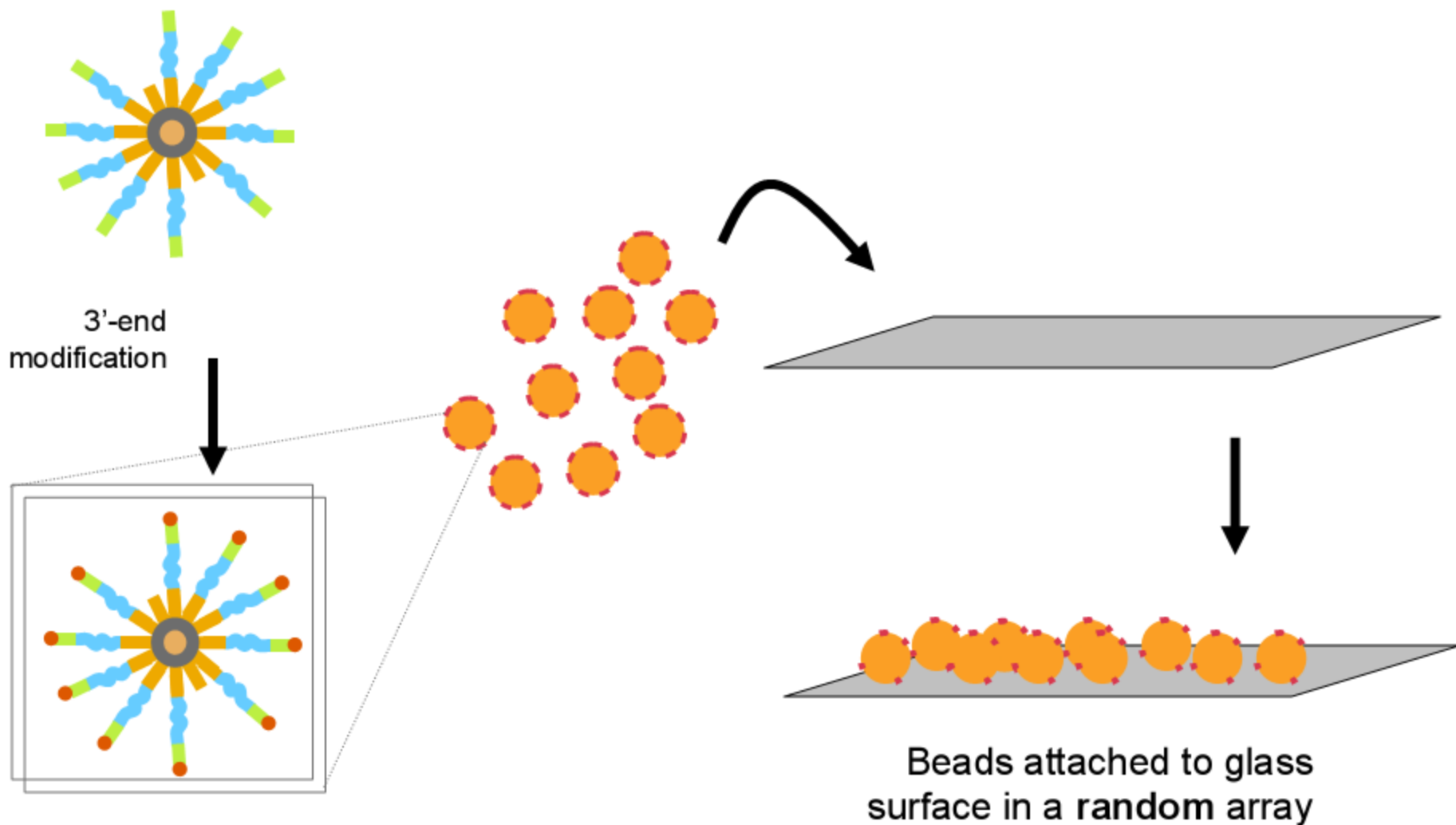


Supernatant
Captured beads with templates

Pellet
Beads with no template

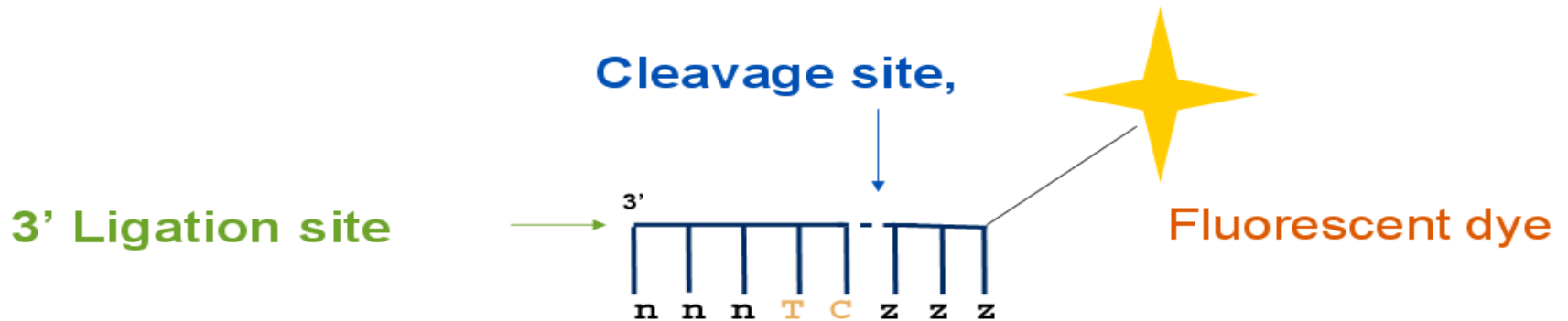


Bead Deposition



Properties of the Probes

Spatial separation among dye, ligation & cleavage sites



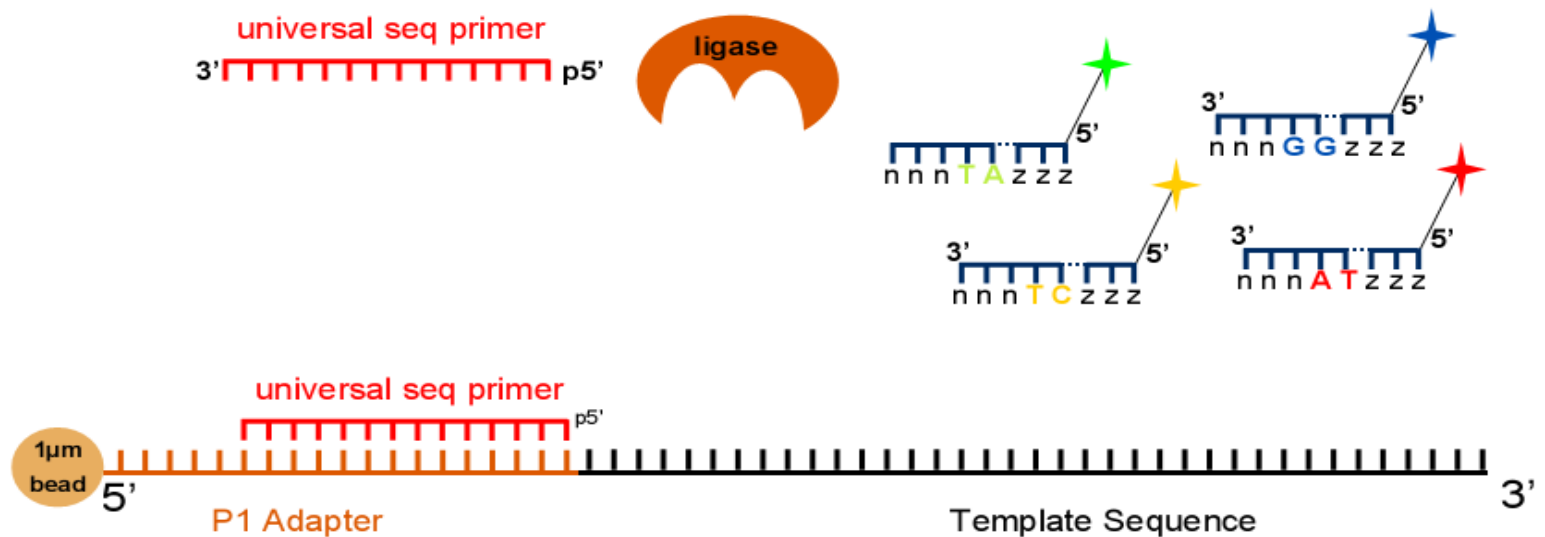
1,024 Octamer Probes (4^5)

4 Dyes, 4 dinucleotides, 256 probes per dye

N= degenerate bases Z= Universal bases

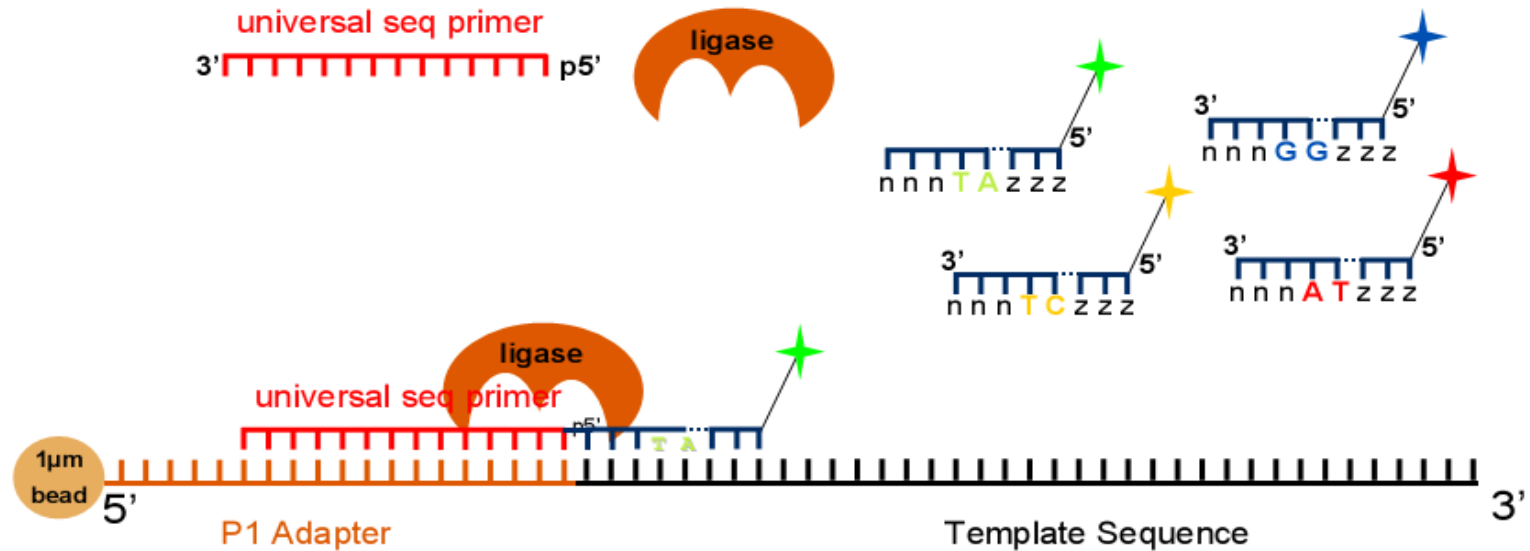
SOLiD Chemistry System 4-color ligation

Ligation reaction

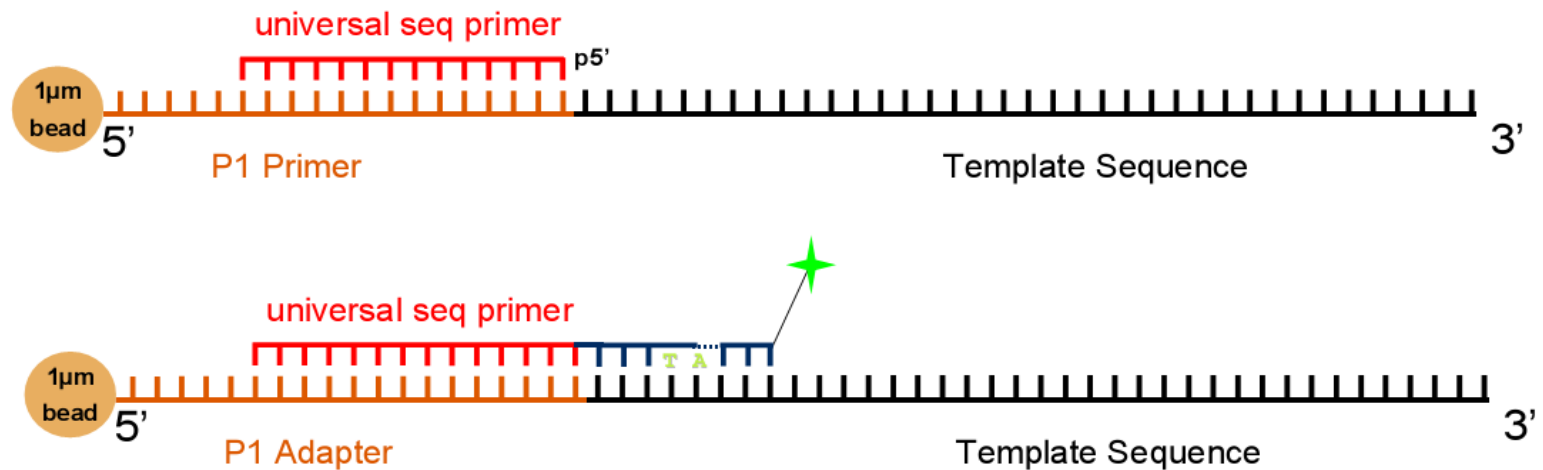


SOLiD Chemistry System 4-color ligation

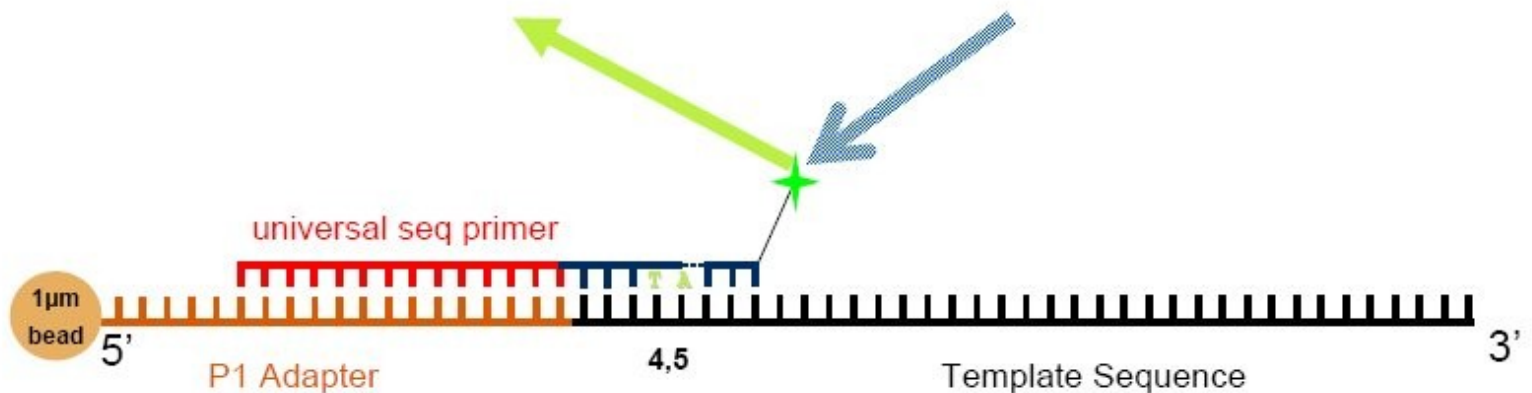
Ligation reaction



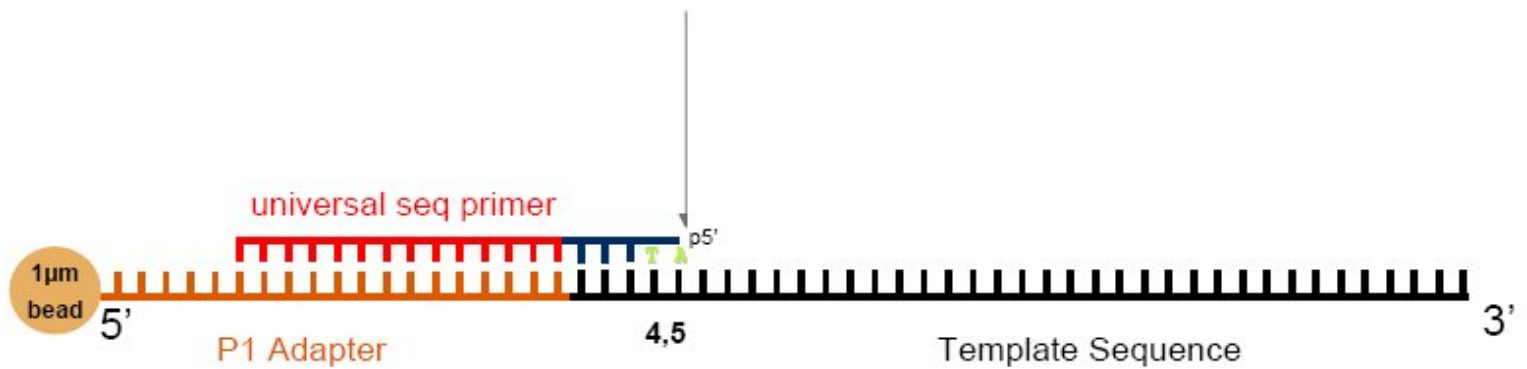
SOLiD Chemistry System 4-color ligation De-Phosphorylation



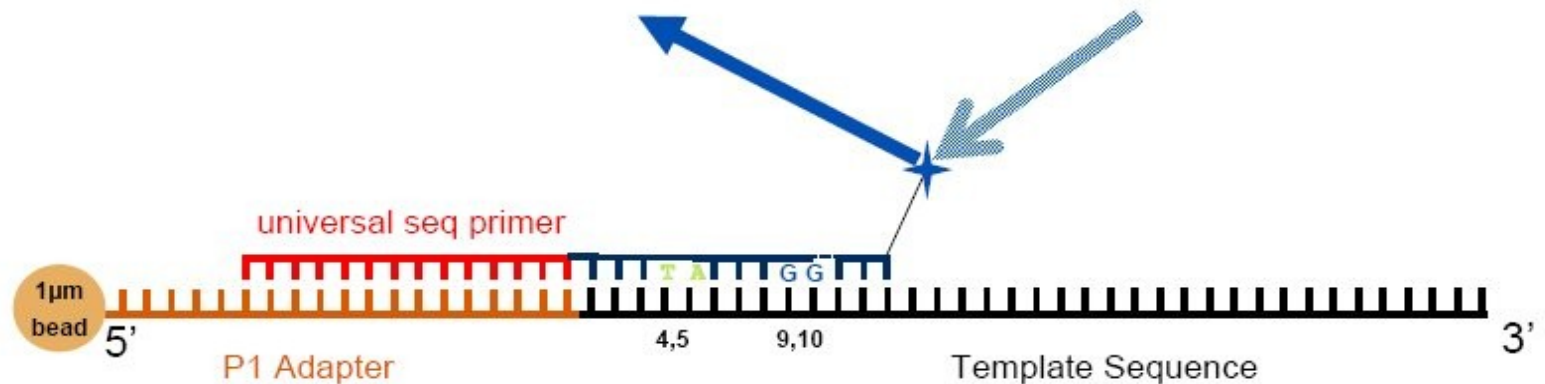
SOLiD Chemistry System 4-color ligation Visualization



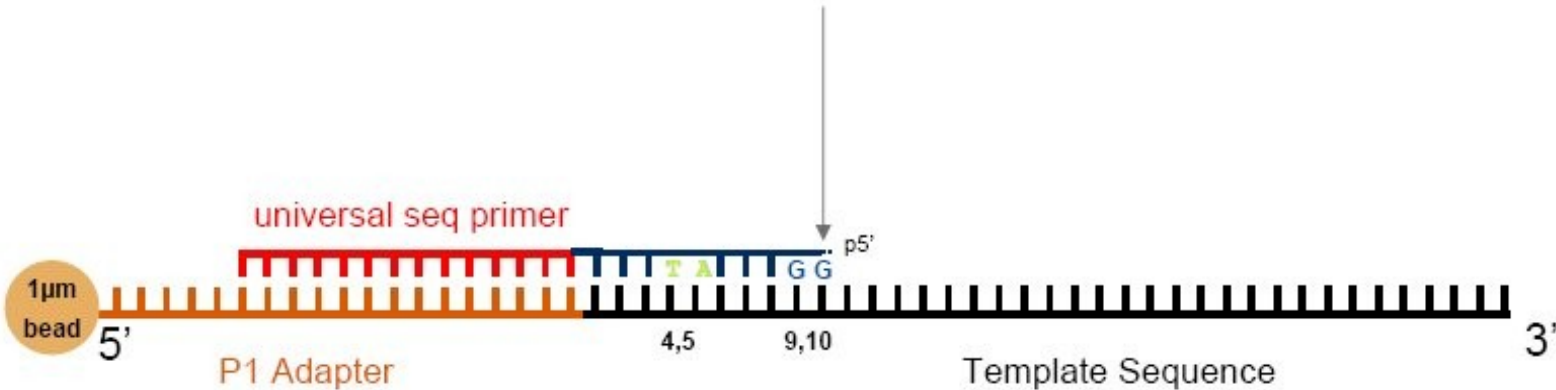
SOLiD Chemistry System 4-color ligation Cleavage



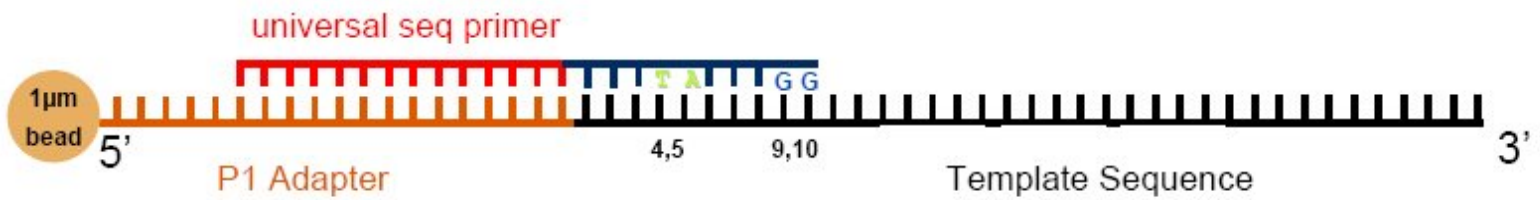
SOLiD Chemistry System 4-color ligation Visualization (2nd cycle)



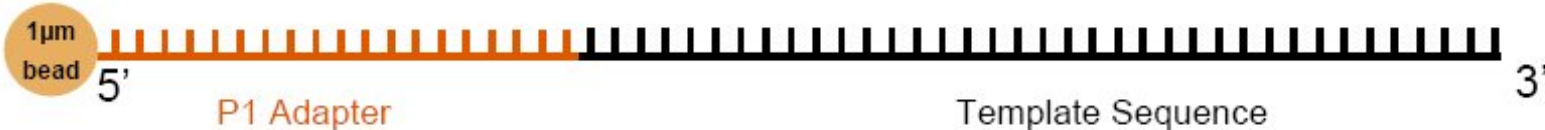
SOLiD Chemistry System 4-color ligation Cleavage (2nd cycle)



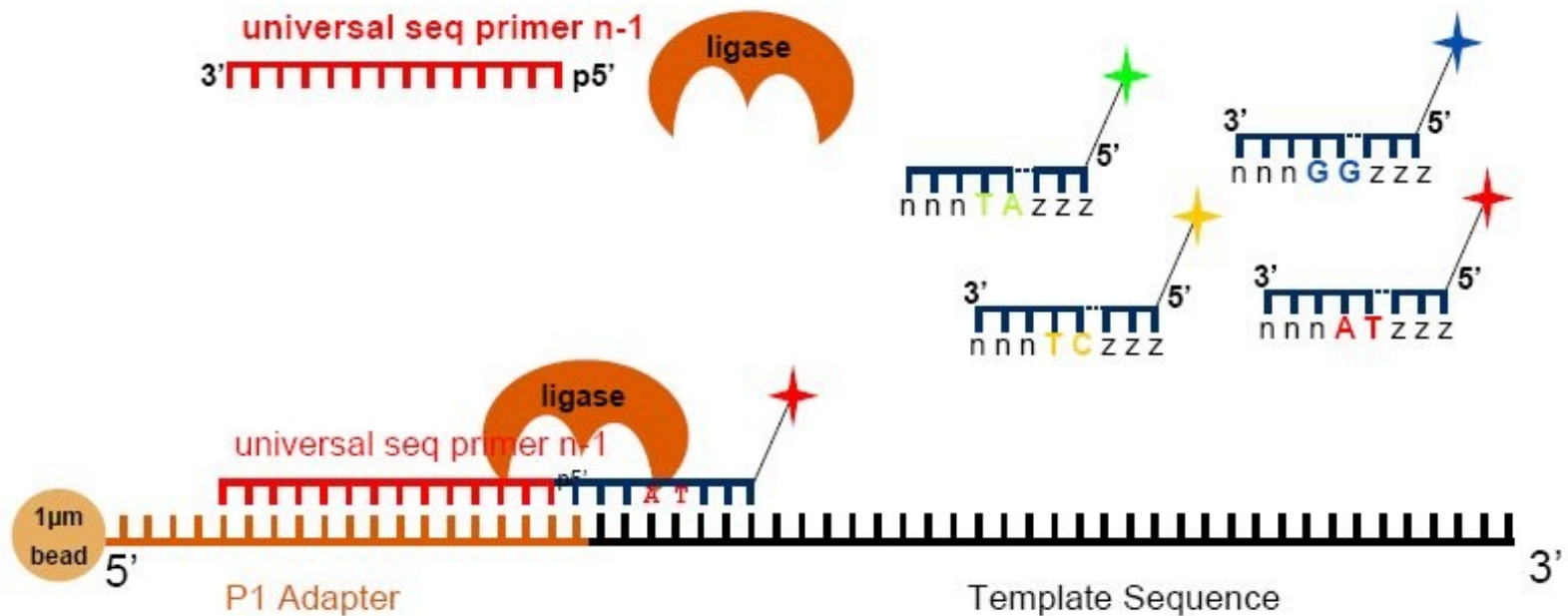
SOLiD Chemistry System 4-color ligation interrogates every 5th base



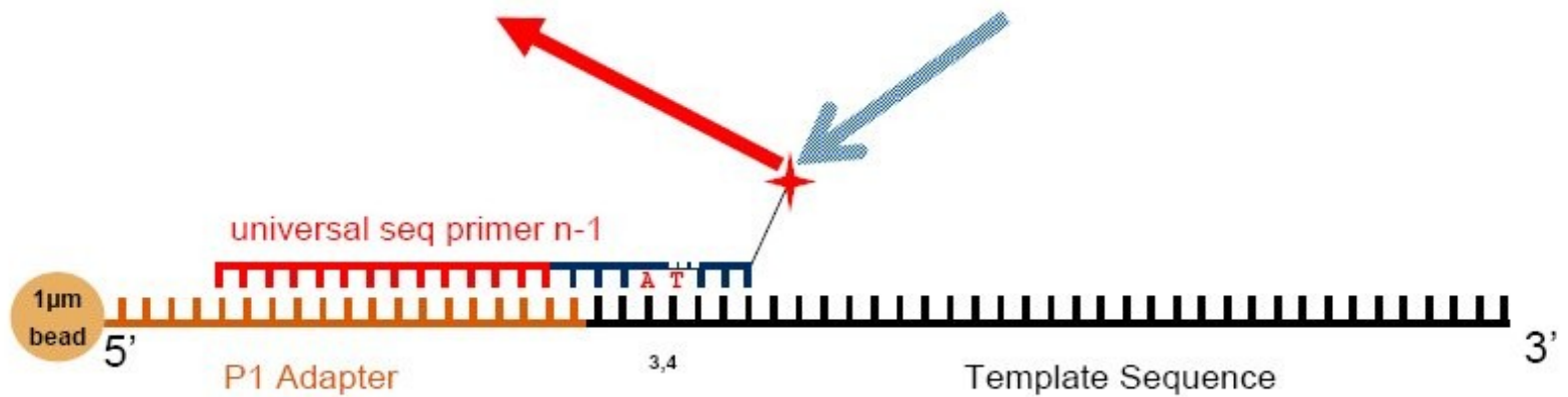
SOLiD Chemistry System 4-color ligation Reset



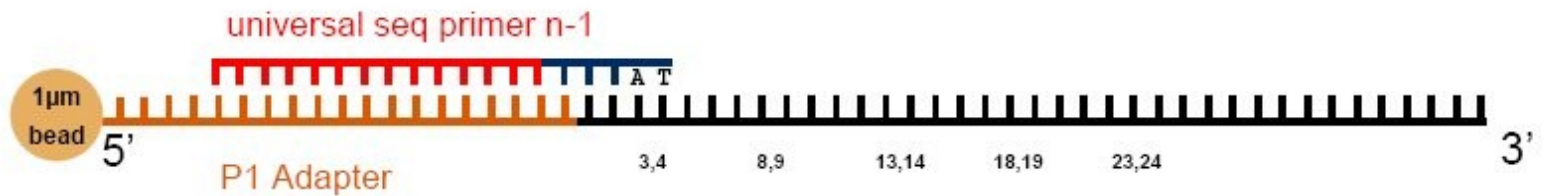
SOLiD Chemistry System 4-color ligation (1st cycle after reset)



SOLiD Chemistry System 4-color ligation (1st cycle after reset)

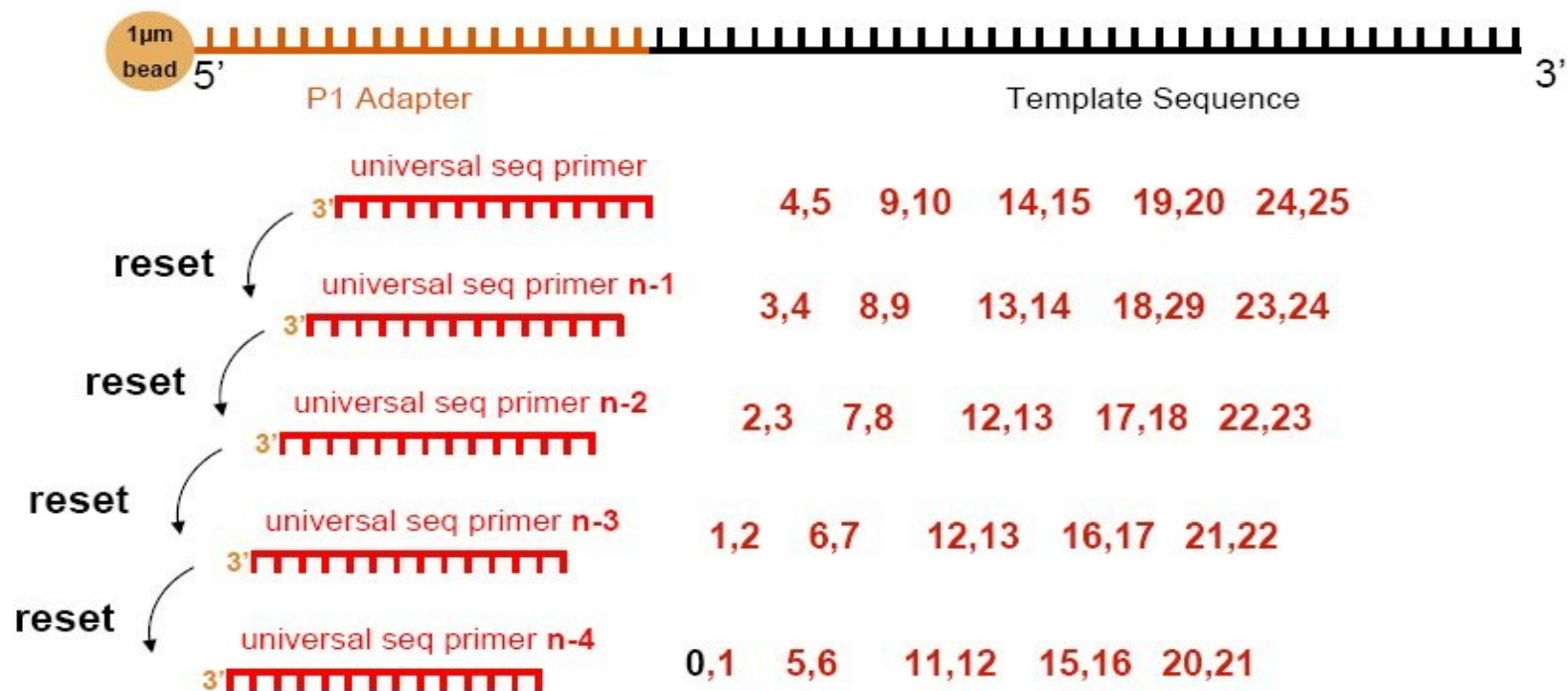


SOLiD Chemistry System 4-color ligation (2nd Round)



Sequential rounds of sequencing

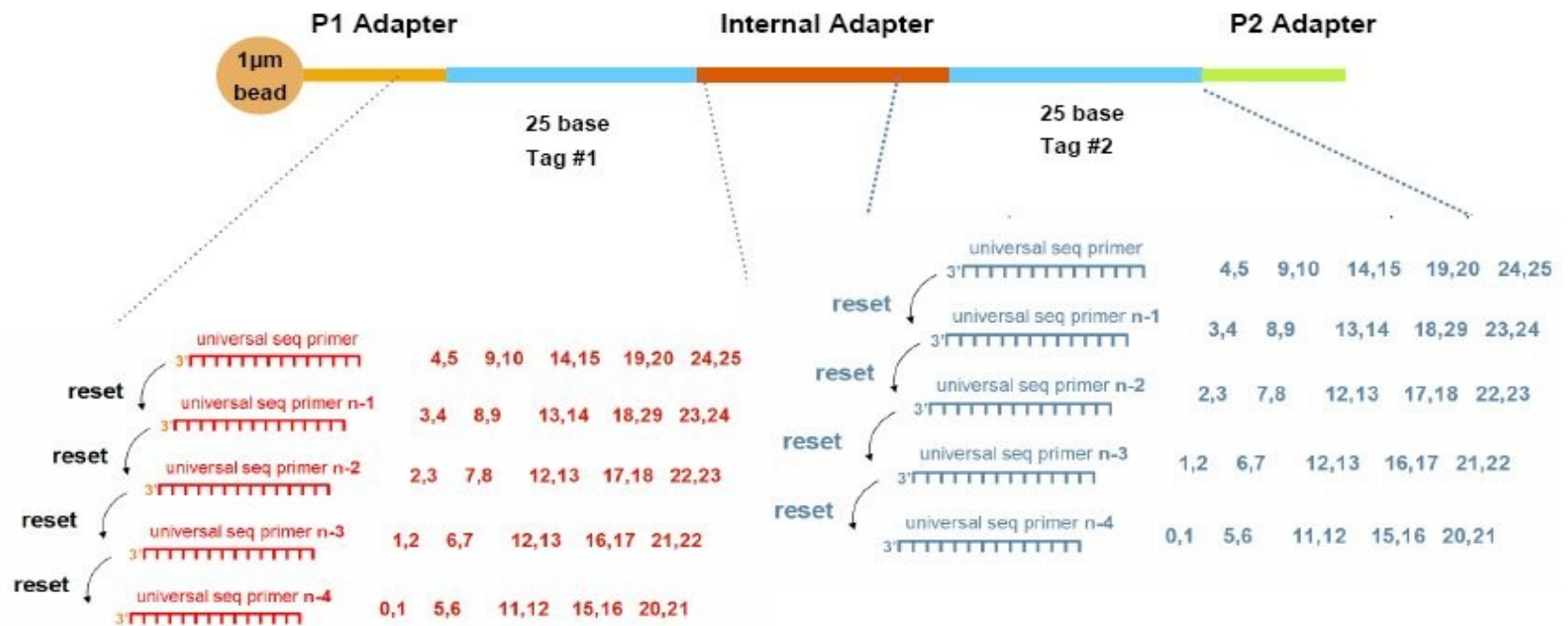
Multiple cycles per round



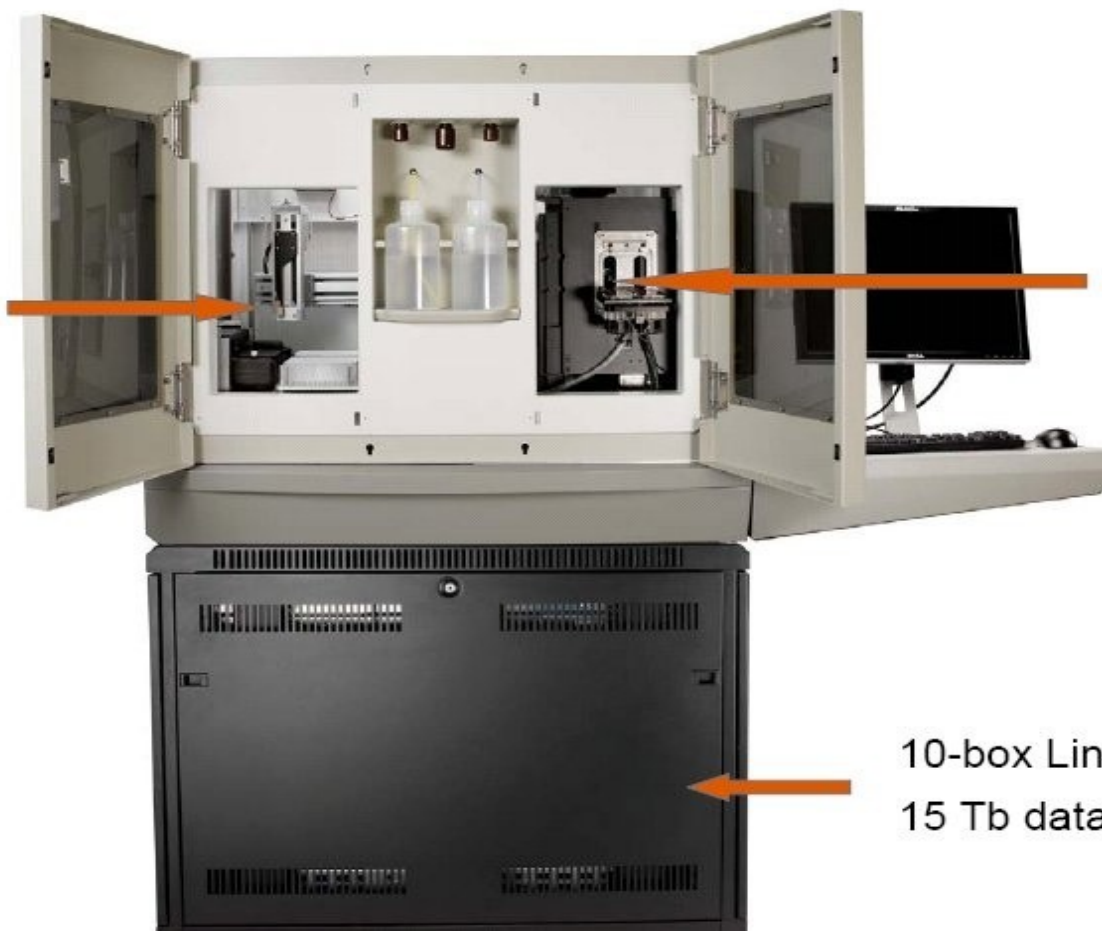
Paired End two sequences generated

Sequential rounds of sequencing

Multiple cycles per round



Reagent
handling



Dual Flow
Cell

10-box Linux Cluster
15 Tb data storage

SOLiD System Arrays: 4 Mega pixel Camera

Full Array



1800 panels

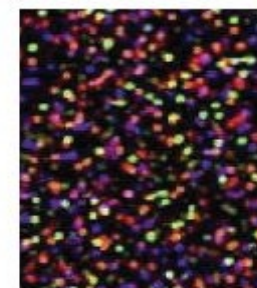
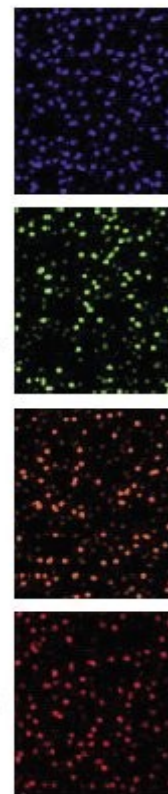
$1/6^{\text{th}}$



Small Array



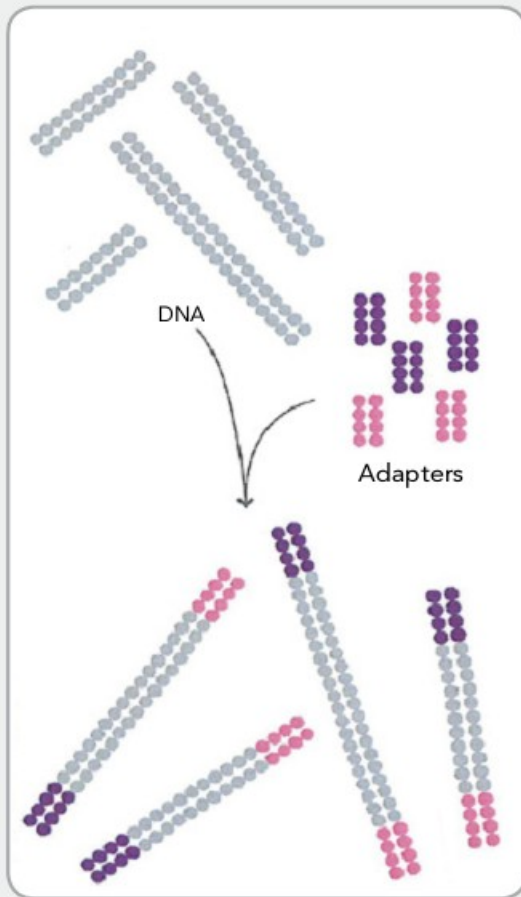
300 panels
~20,000 beads/panel



4-Color Overlay

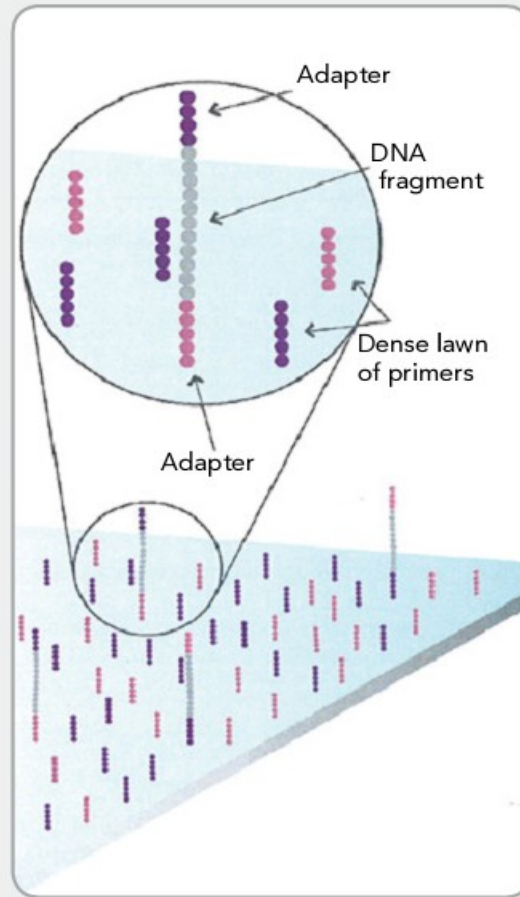
Solexa (2007)

1. PREPARE GENOMIC DNA SAMPLE



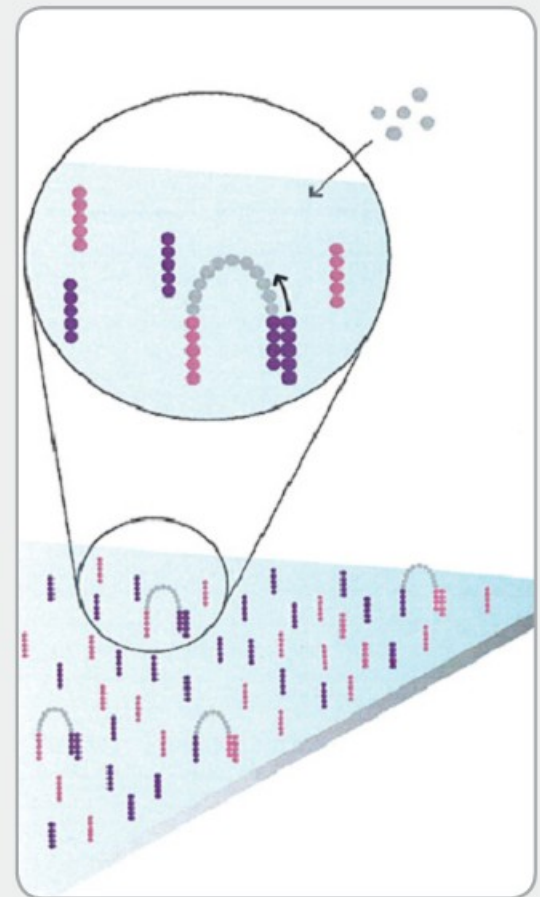
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



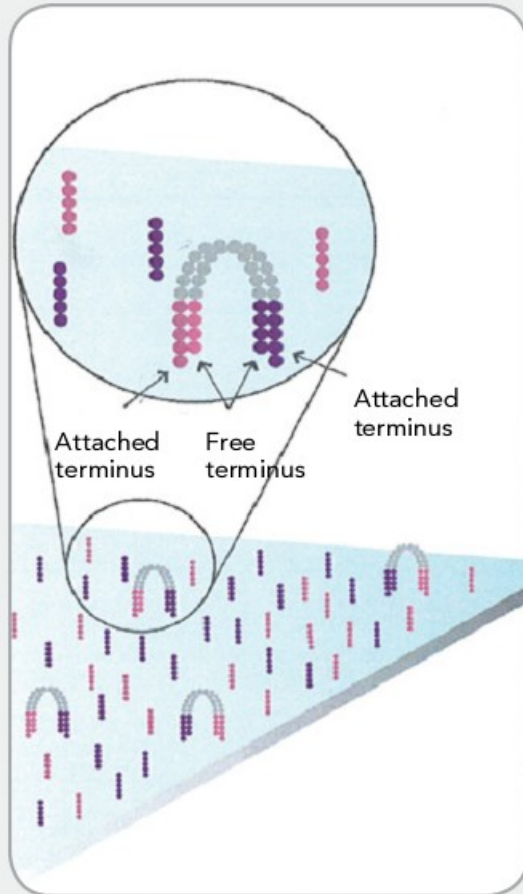
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION



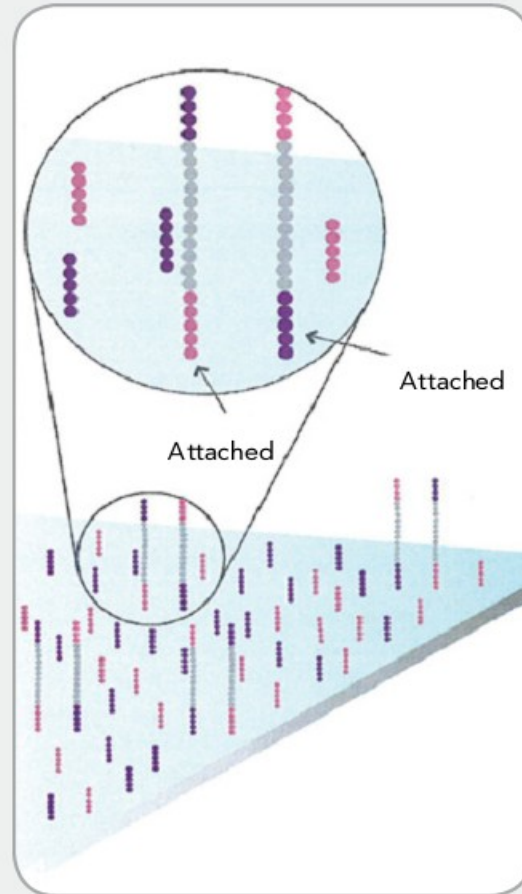
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4. FRAGMENTS BECOME DOUBLE STRANDED



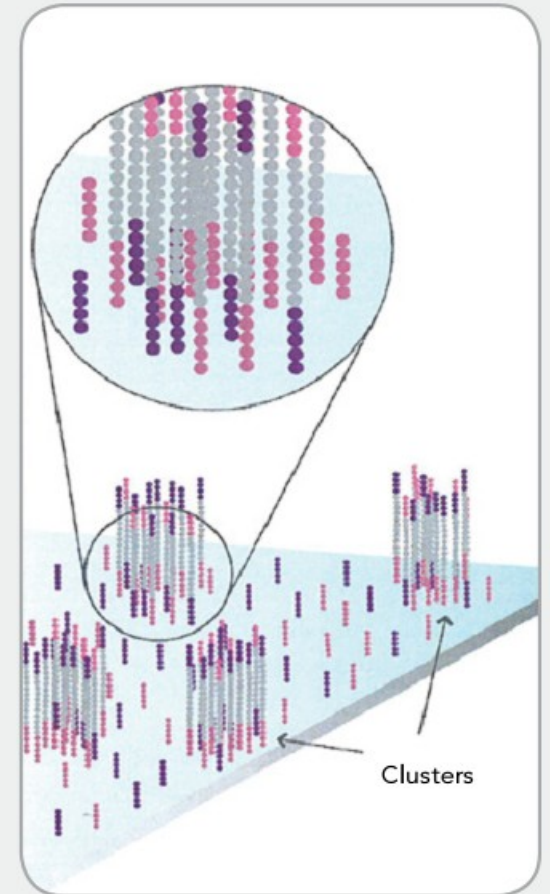
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



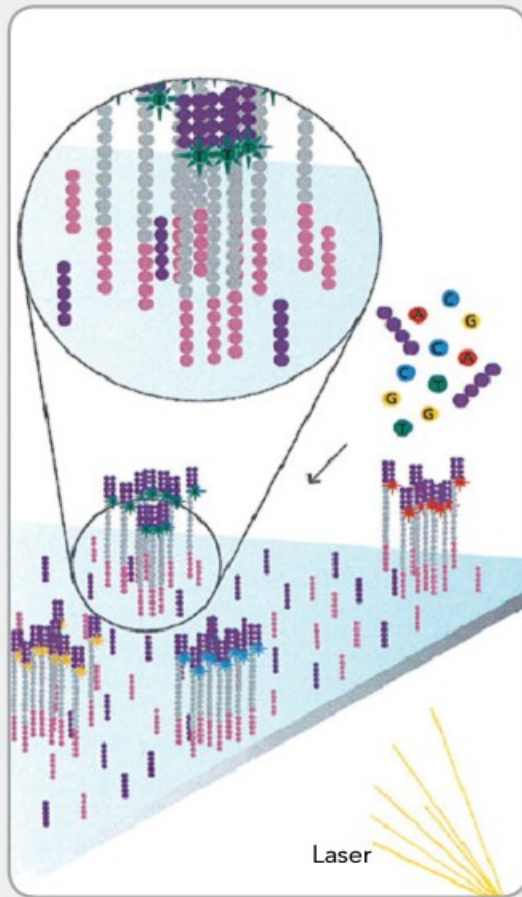
Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION



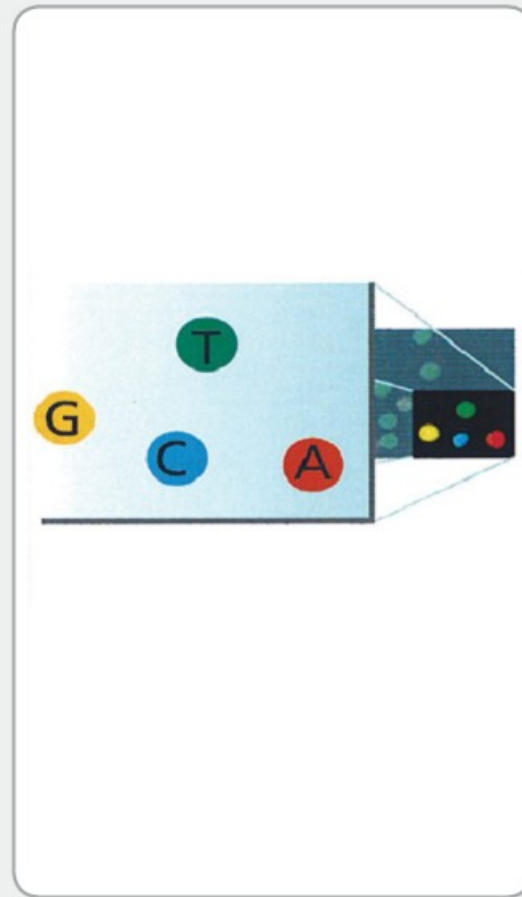
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

7. DETERMINE FIRST BASE



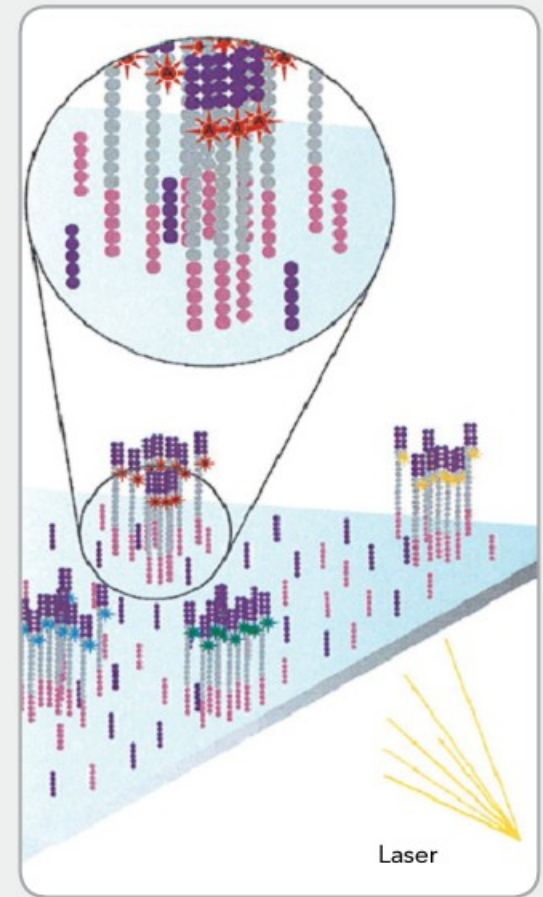
First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8. IMAGE FIRST BASE



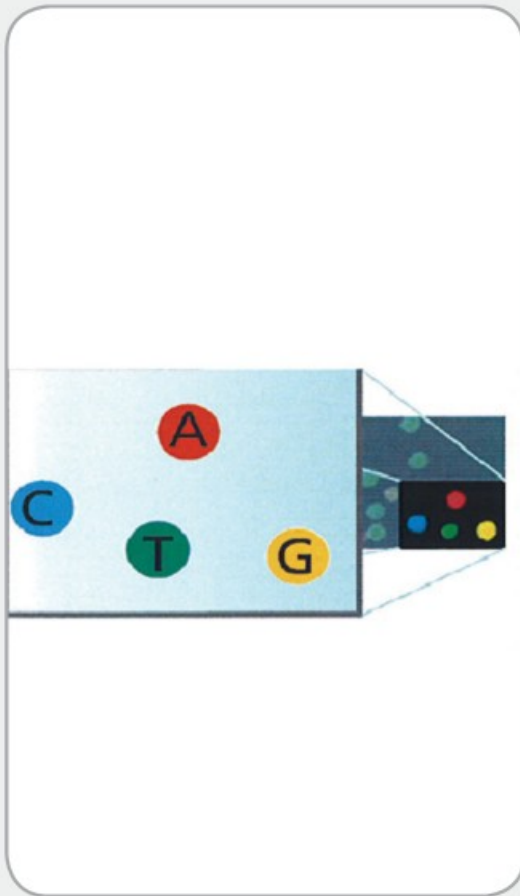
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

9. DETERMINE SECOND BASE



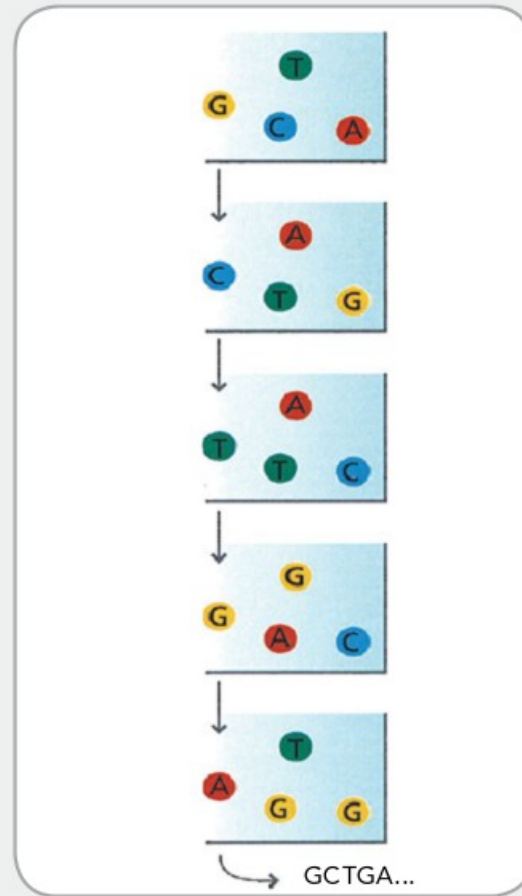
Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

10. IMAGE SECOND CHEMISTRY CYCLE



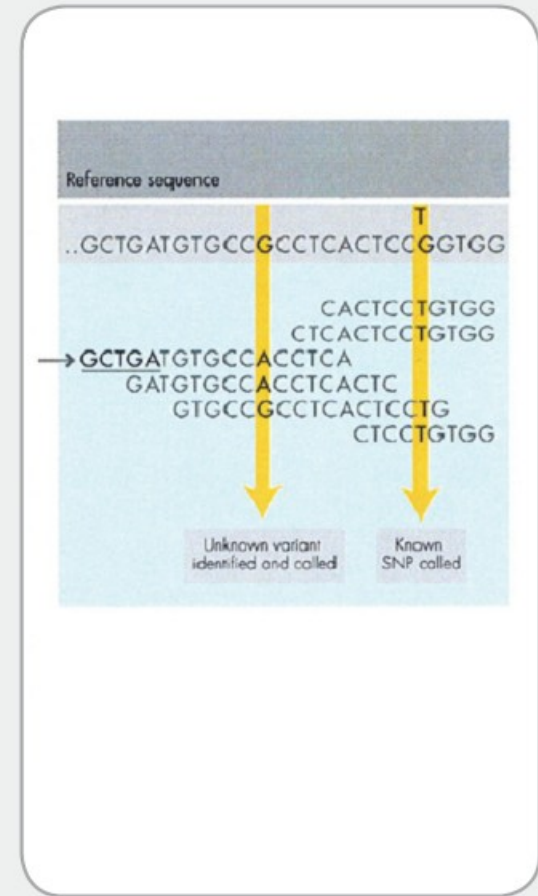
After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

11. SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES



Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

12. ALIGN DATA

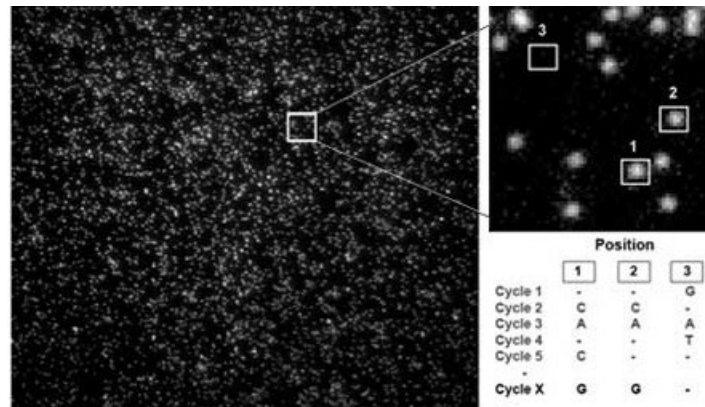
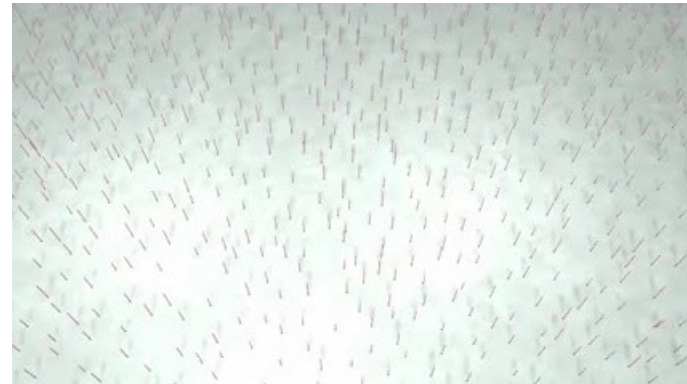
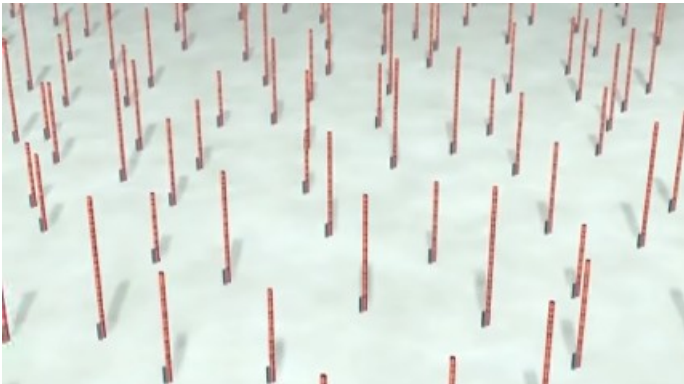


Align data, compare to a reference, and identify sequence differences.

HELICOS (2008)



True Single Molecule Sequencing (tSMS)

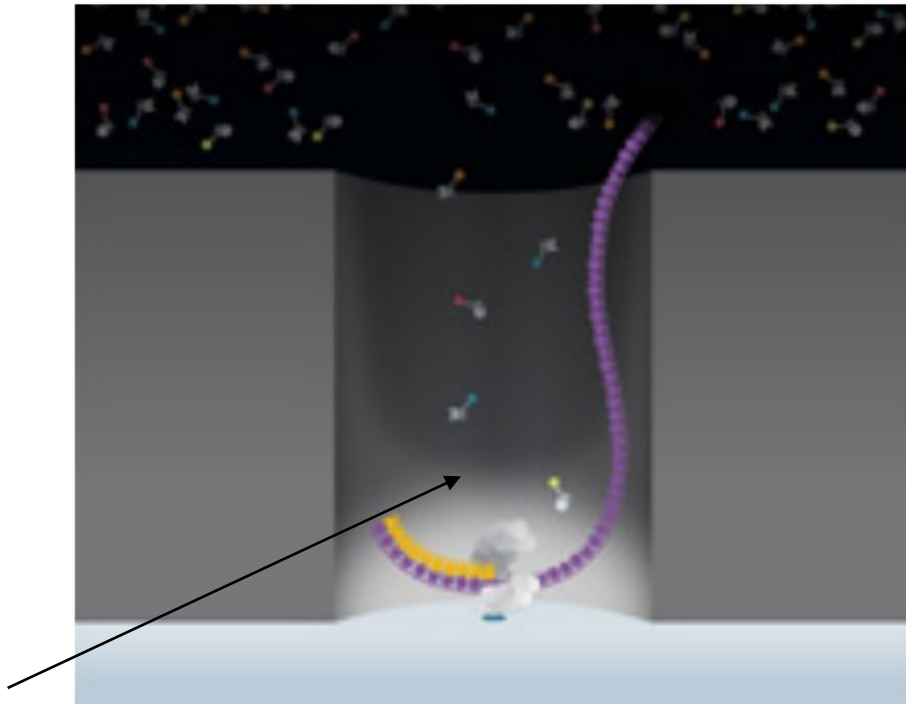


Coming soon: 15-minute, \$100 human genome sequencing

Single Molecule Real-Time (SMRT)

Pacific Biosciences

20 zeptolitru



Další technologie

- Nanopore sekvenování
- Mikroelektroforéza
- Sekvenování na bázi microarray