

# Sekvenční analýza

**doc. RNDr. Milan Bartoš, Ph.D.**

**bartosm@vfu.cz**

**Přírodovědecká fakulta MU, 2013**

# ***Obsah přednášky***

- 1) Pojem sekvenování**
- 2) Maxamovo-Gilbertovo sekvenování**
- 3) Sangerova metoda s využitím fluoroforů**
- 4) Pyrosekvenování**
- 5) Nové přístupy k sekvenování**
- 6) Sekvenování RNA**
- 7) Sekvenování proteinů**



# *Doporučená literatura*

[www.farmakogenomika.cz](http://www.farmakogenomika.cz)

# ***Sekvenování***

## **Rozhodující metoda pro stanovení nukleotidových sekvencí**

- **Konečná fáze procesu individualizace jednotlivých izolátů**
- **Metoda je pro většinu mikrobiologických aplikací příliš přesná**

# ***Metody sekvenování nukleových kyselin***

**Chemická metoda sekvenování**  
(Maxamovo-Gilbertovo sekvencování)

**Enzymová metoda sekvenování**  
(Sangerovo sekvenování)

**Pyrosekvenování**

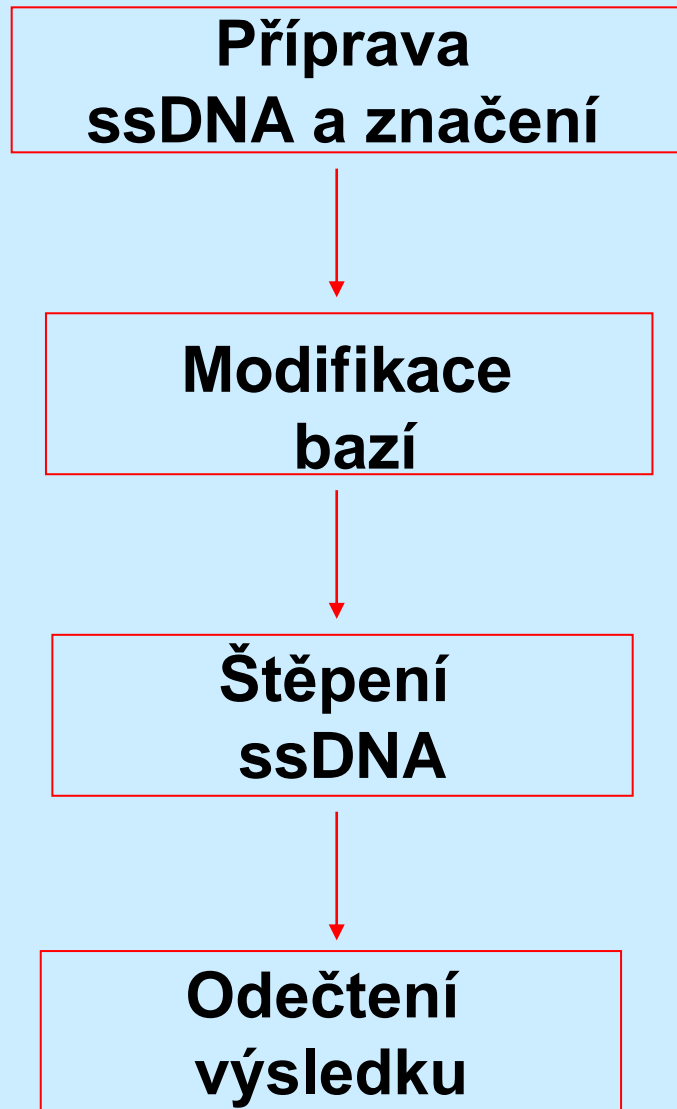
# ***Chemická metoda sekvenování*** ***(Maxamovo-Gilbertovo sekvenování)***

Podstatou je specifické štěpení molekuly **ssDNA** po modifikaci jednotlivých bází chemickými činidly s následnou elektroforetickou detekcí v denaturujícím polyakrylamidovém gelu.

Chemická činidla jsou **specifická** pro modifikaci určitých bází:

<b>G</b>	– DMS
<b>A+G</b>	– piperidin
<b>C+T</b>	– hydrazin
<b>C</b>	– hydrazin + NaCl

# *Chemická metoda sekvenování*



# Chemická metoda sekvenování

Příprava  
ssDNA a značení

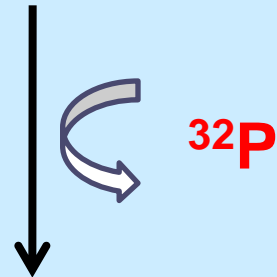


Modifikace  
bází

Asymetrická PCR

Využití vazby  
biotinylovaného primeru

Štěpení  
ssDNA



Odečtení  
výsledku





# Chemická metoda sekvenování

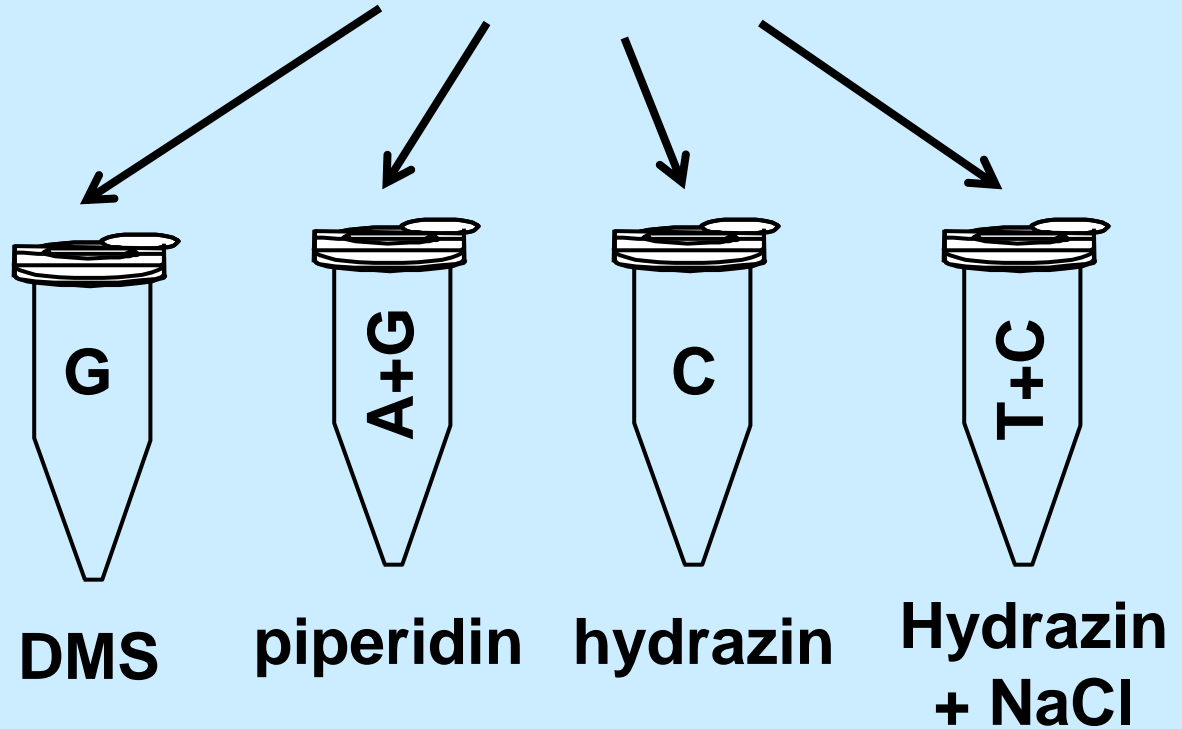
Příprava  
ssDNA a značení

Modifikace  
bází

Štěpení  
ssDNA

Odečtení  
výsledku

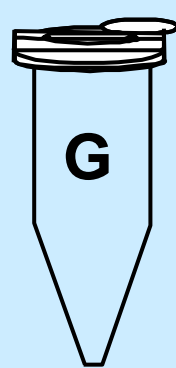
$^{32}\text{P}$  -GATCAGG - 3'



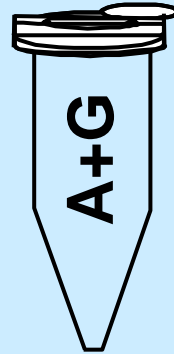
# Chemická metoda sekvenování

<sup>32</sup>P -GATCAGG - 3'

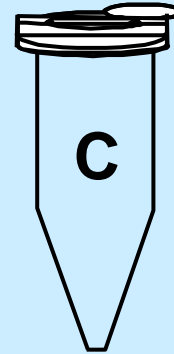
Příprava  
ssDNA a značení



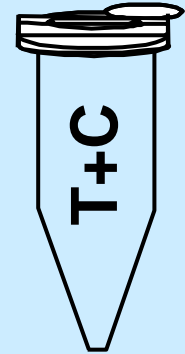
DMS



piperidin



hydrazin



Hydrazin  
+ NaCl

Modifikace  
bází

Štěpení  
ssDNA

Odečtení  
výsledku

Štěpení piperidinem při vysoké teplotě

<sup>32</sup>P -GATCAGG/G

<sup>32</sup>P -G/ATCA/G/G

<sup>32</sup>P -GATC/AGG

<sup>32</sup>P -GAT/C/AGG

**<sup>32</sup>P** -GATCAGG - 3'

**G**

DMS



**<sup>32</sup>P** -GATCAG  
**<sup>32</sup>P** -GATCAGG

**A+G**

piperidin



**<sup>32</sup>P** - GA  
**<sup>32</sup>P** - GATCA  
**<sup>32</sup>P** - GATCAG  
**<sup>32</sup>P** - GATCAGG

**T+C**

hydrazin



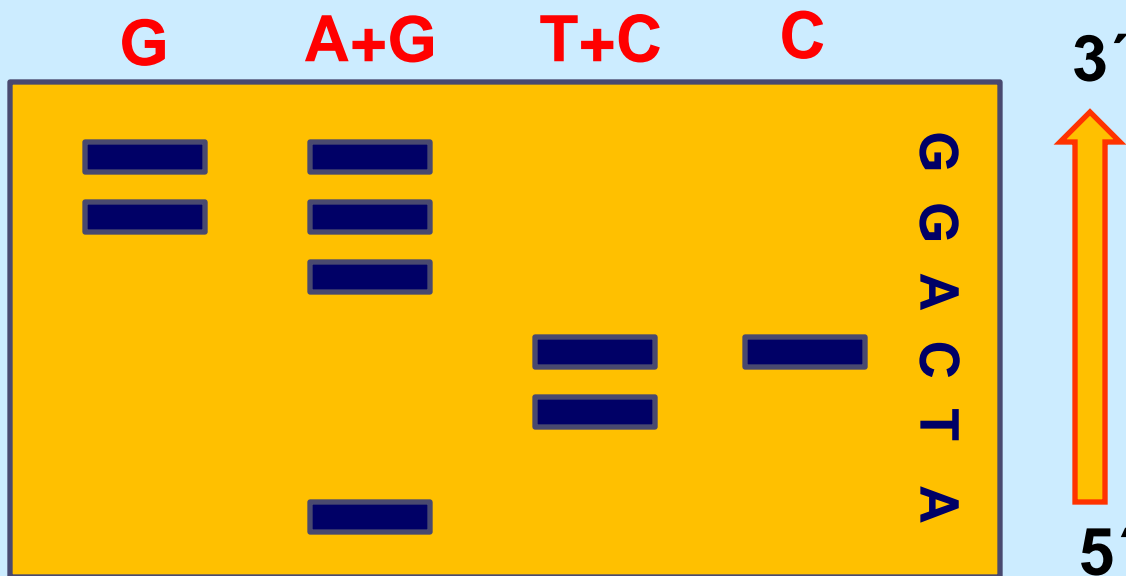
**<sup>32</sup>P** - GAT  
**<sup>32</sup>P** - GATC

**C**

Hydrazin  
+ NaCl

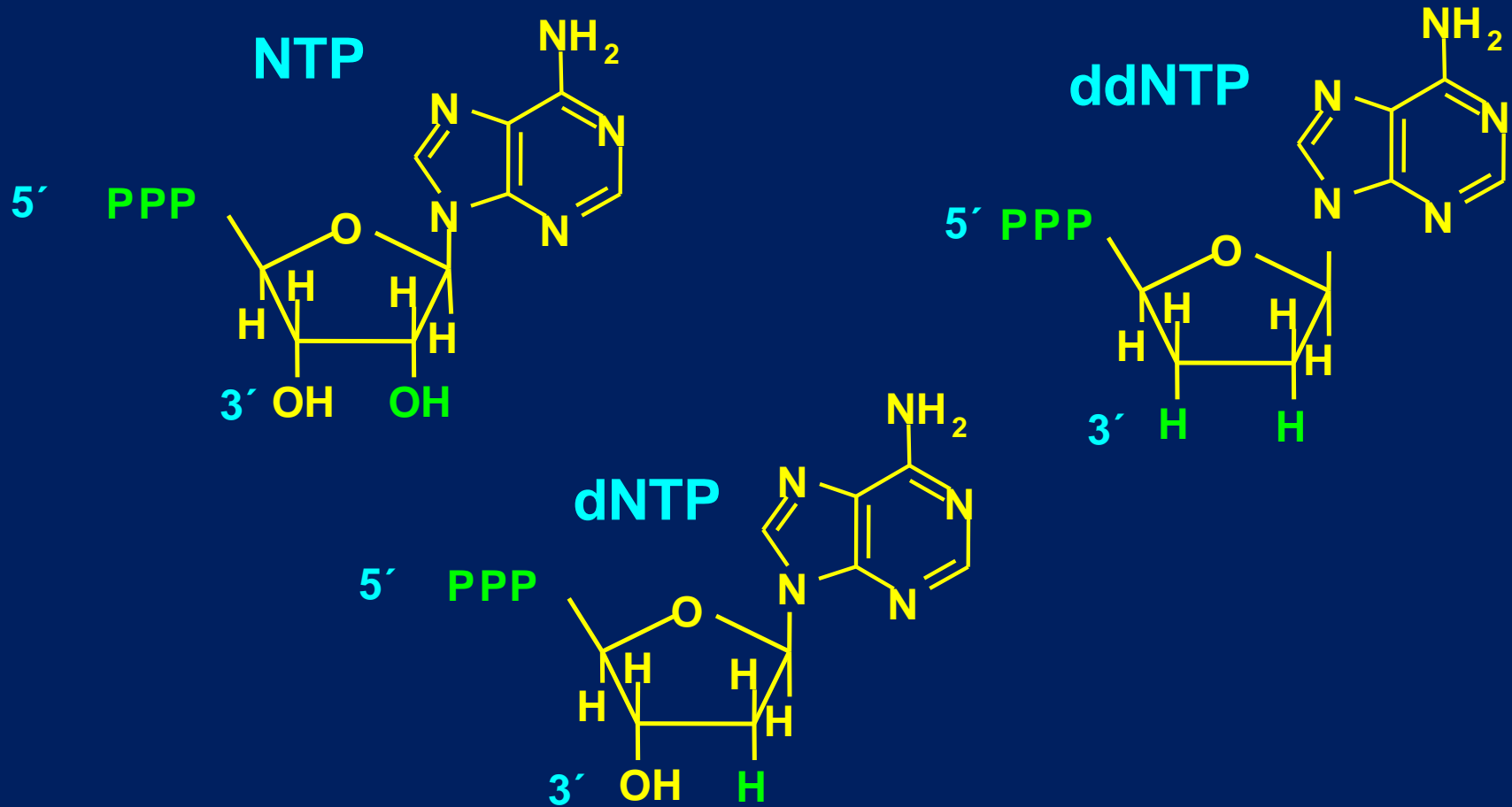


**<sup>32</sup>P** - GATC

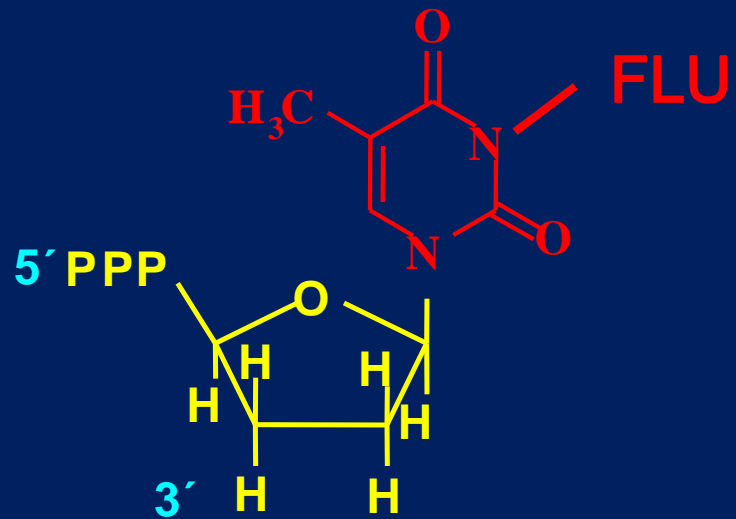
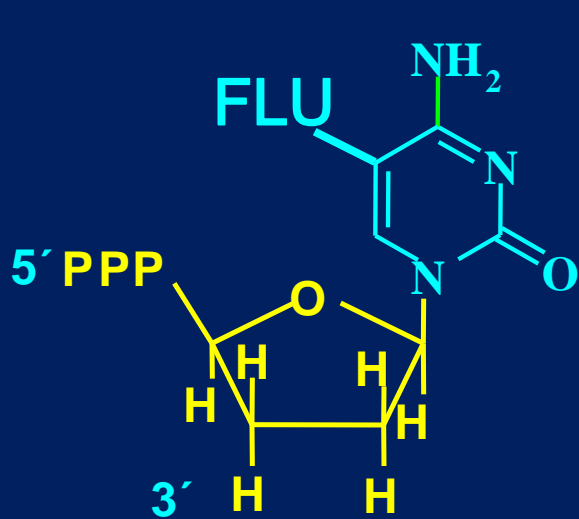
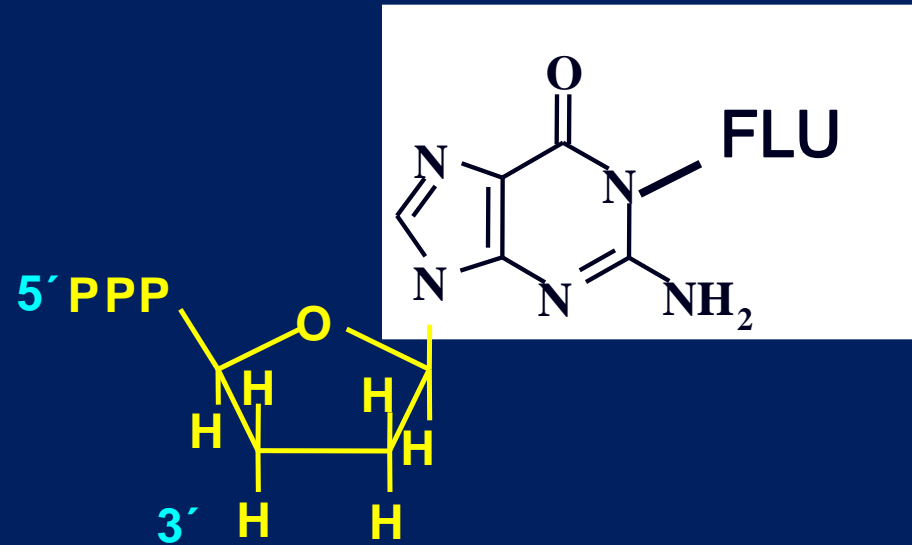
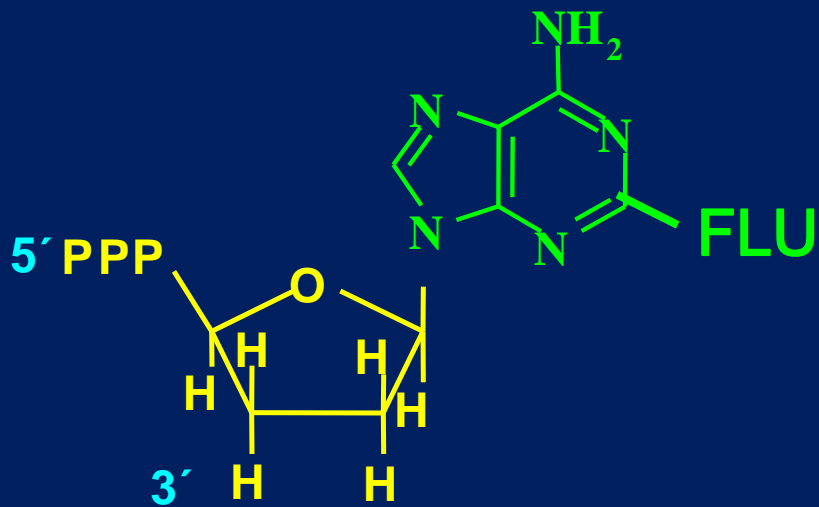


# Sangerova metoda

## dideoxyterminátory



# Dideoxyterminátory



# *Průběh sekvenování*

1. denaturace (92-96°C)



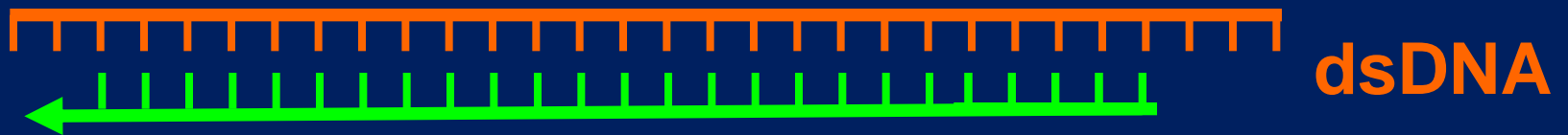
2. annealing (45-72°C)

3. extenze (72°C)



# *Průběh sekvenování*

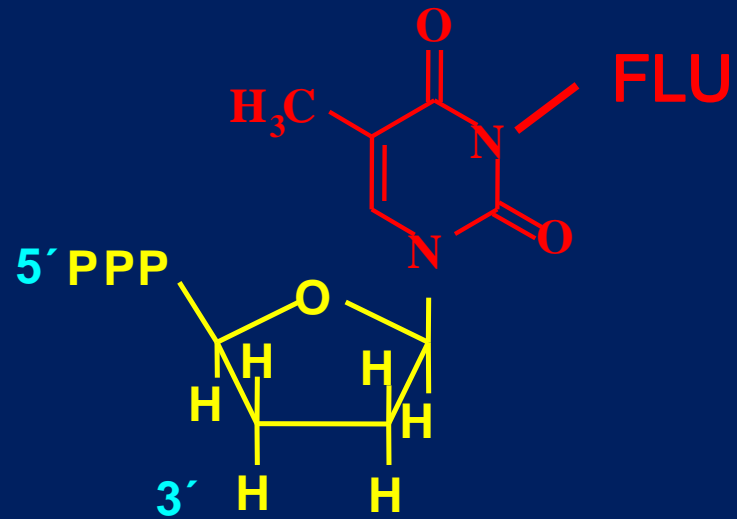
1. denaturace (92-96°C)



2. annealing (45-72°C)

3. extenze (72°C)

# Zařazování ddTTP



TTTTTGTCAAATCGGTGTAAACGGT

TTTTTGTCAAATCGGTGT

TTTTTGTCAAATCGGT

TTTTTGTCAAA

ATGCTACGTTTAGCCACATTTGCCATATGAACCT





# Zařazování dalších ddNTP

TTTTTGTCAAATCGGTGTA

TTTTTGTCAA

TTTTTGTCAA

TTTTTGTCA

ATGCTACGTTTAGCCACATTTGCCATATGAACCT



TTTTTGTCAAATCGGTGTAAAC

TTTTTGTCAAATCC

TTTTTGTCAAATC

TTTTTGTCA

ATGCTACGTTTAGCCACATTTGCCATATGAACCT

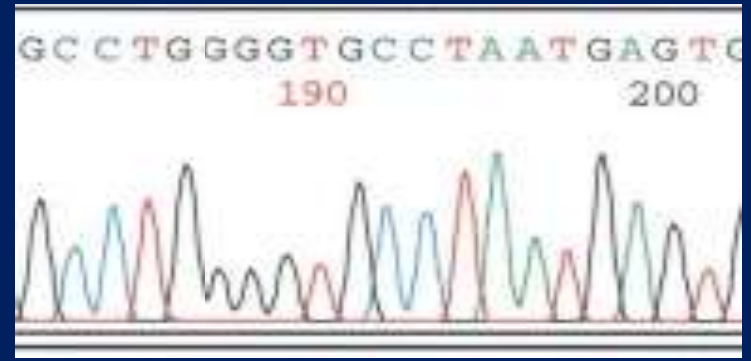
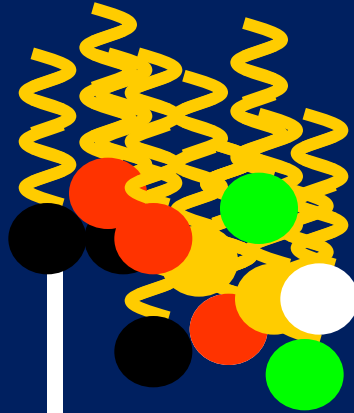




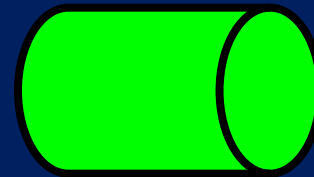
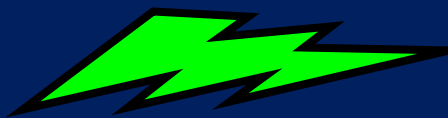
# *Následuje rozdělení fragmentů*



# Následuje rozdělení fragmentů

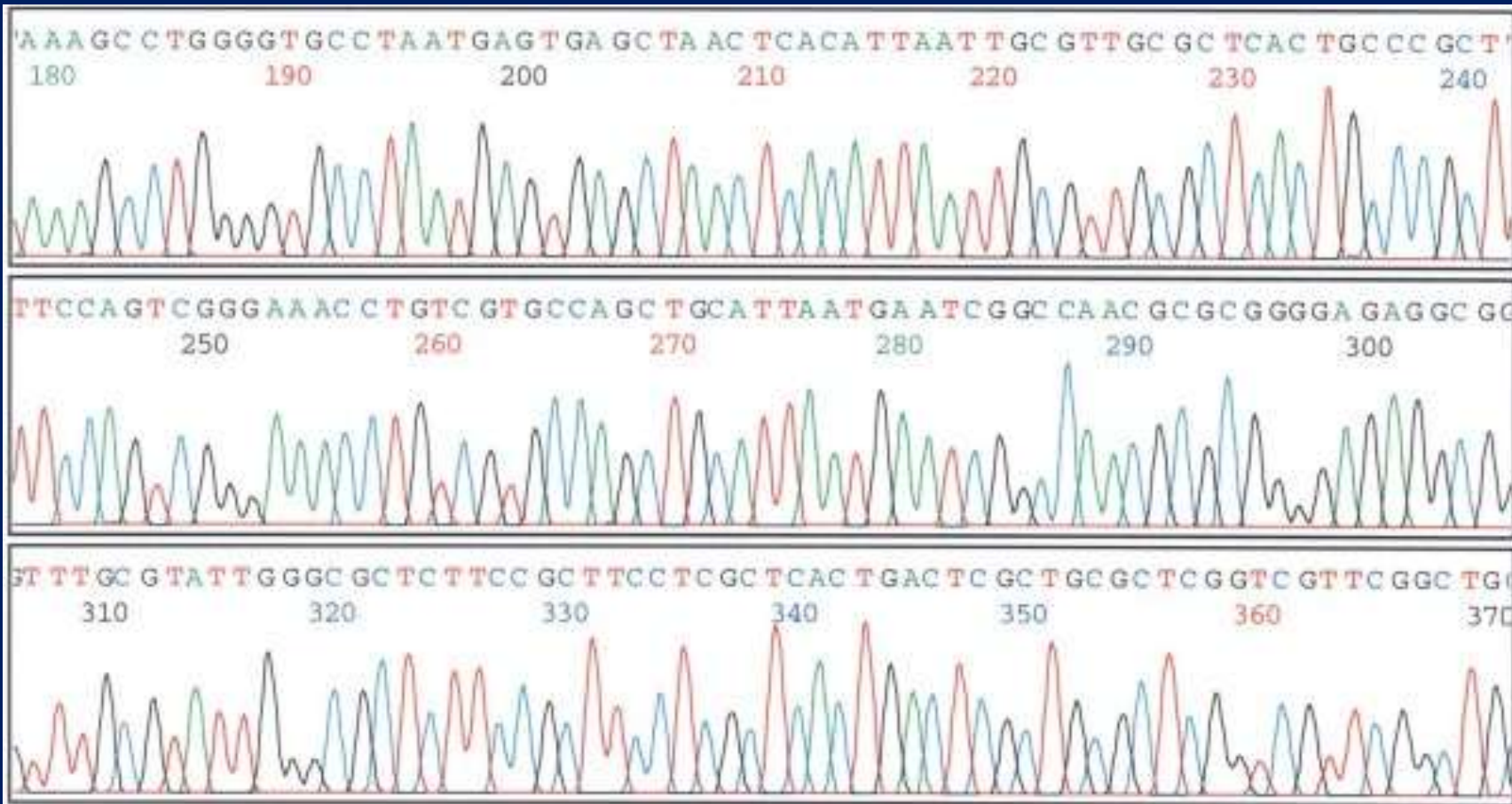


LASER



DETEKTOR

# Sekvenování - záznam



# *Kapilární gelová elektroforéza*



- **rozdělí produkty sekvenování podle velikosti**
- **detekuje fluorofory laserem**

# Úkol



V praktickém cvičení zařad'te na základě výsledků sekvenování izoláty bakterií čeledi *Pasteurellaceae*



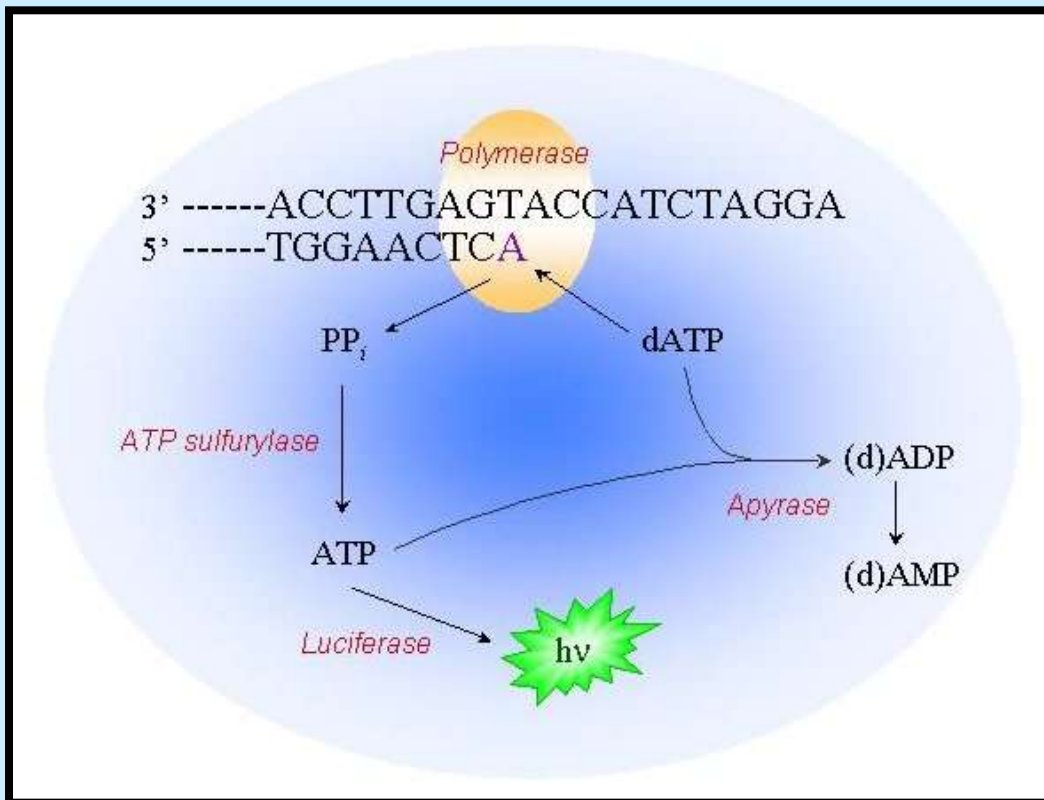
***Prohlédněte si animaci na***

**[http://wwwnc.cdc.gov/eid/article/13/2/06-1032\\_article.htm](http://wwwnc.cdc.gov/eid/article/13/2/06-1032_article.htm)**



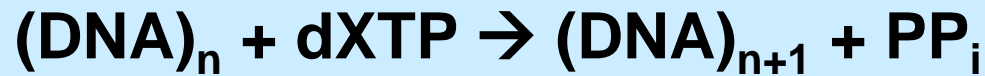
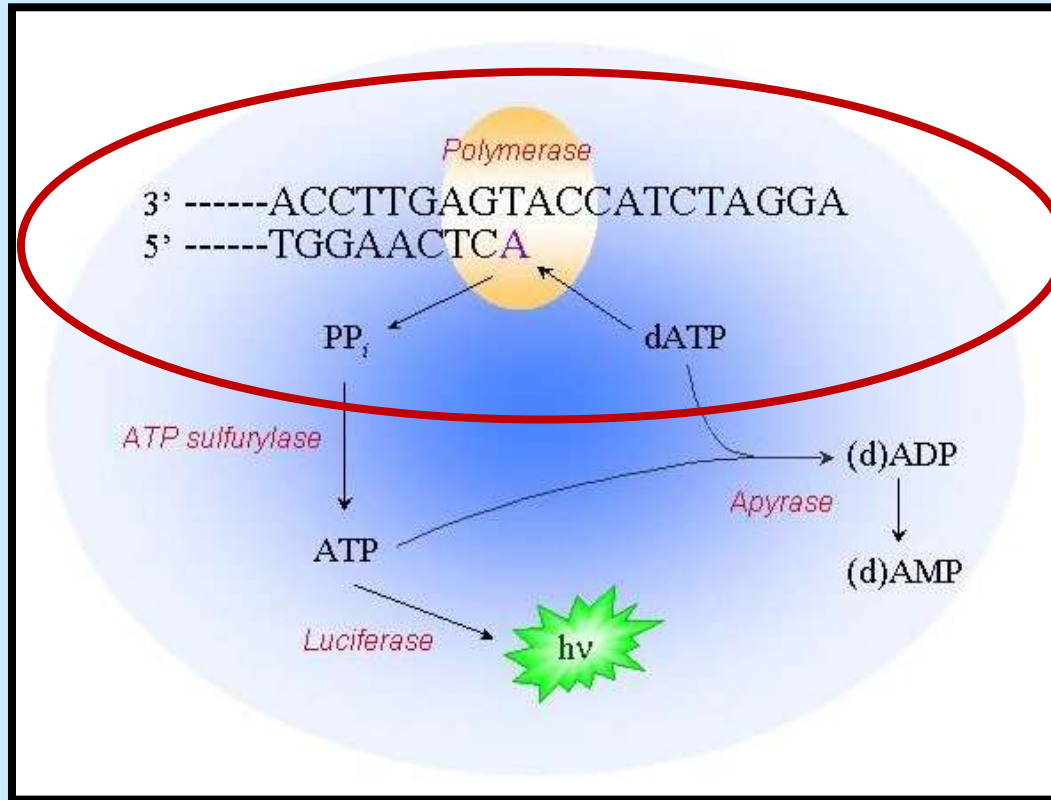
# Pyrosekvenování

- Metoda popsaná Mostafa Ronaghi et al. v roce 1990
- Určena pro krátké úseky DNA, SNP a úseky metylované



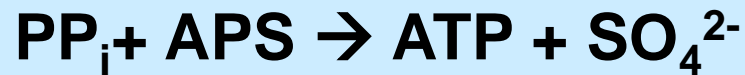
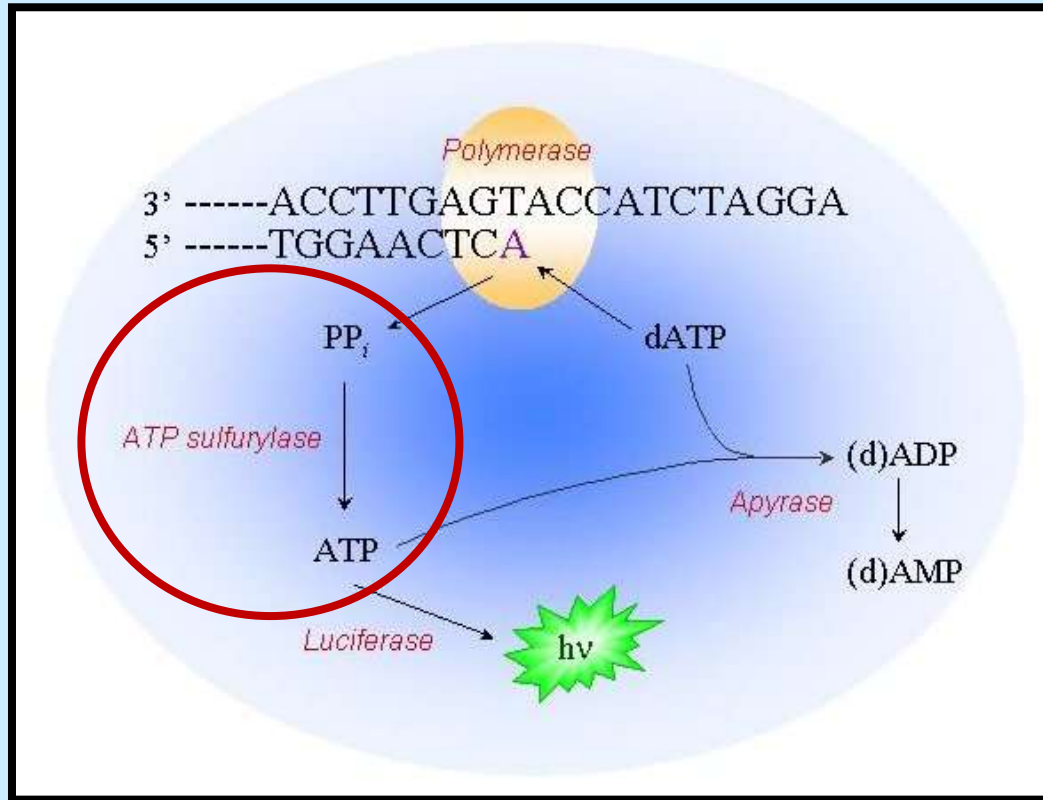
- 4 enzymy
- velmi přesná
- reakce se záhy „zahltí“

# Průběh pyrosekvenování



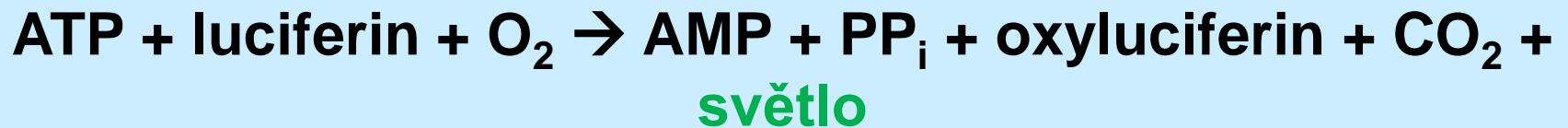
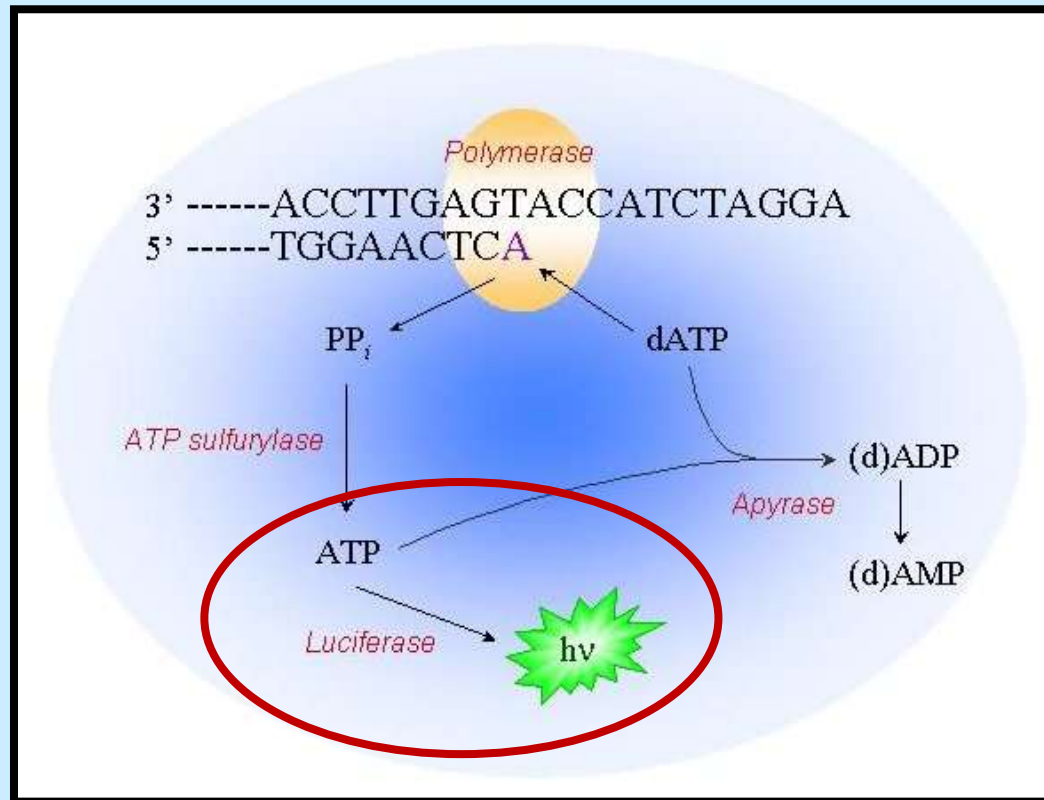
DNA polymeráza

# Průběh pyrosekvenování



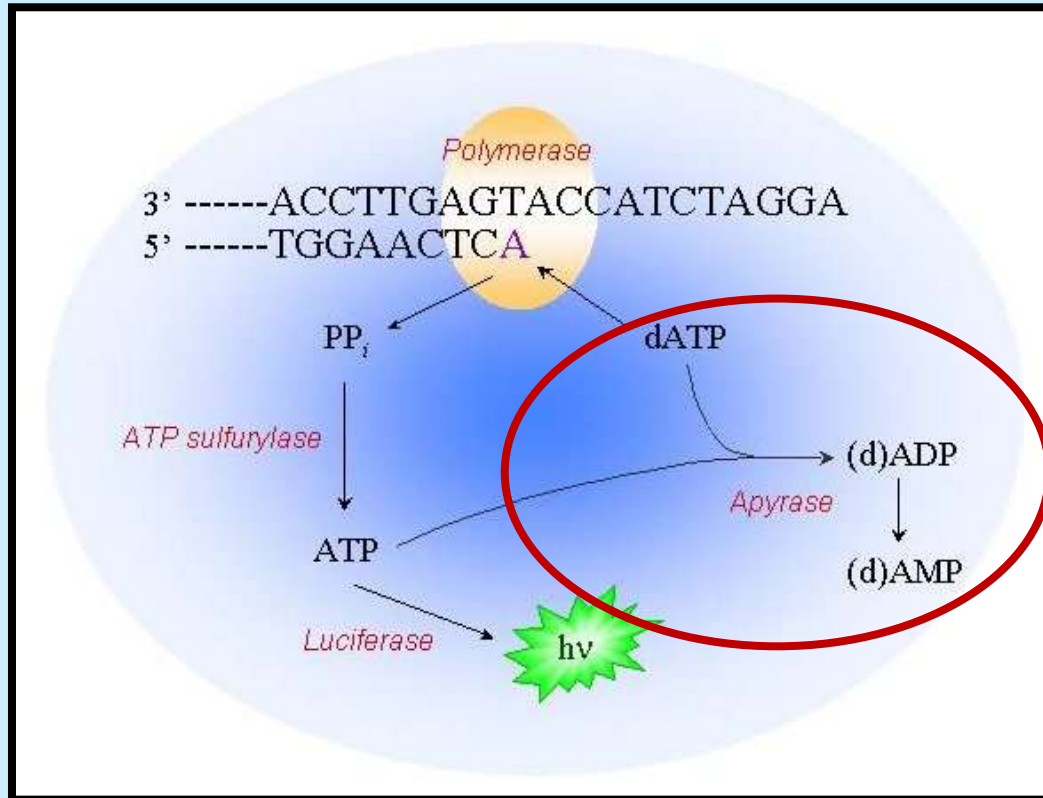
ATP sulfuryláza

# Průběh pyrosekvenování



luciferáza

# Průběh pyrosekvenování



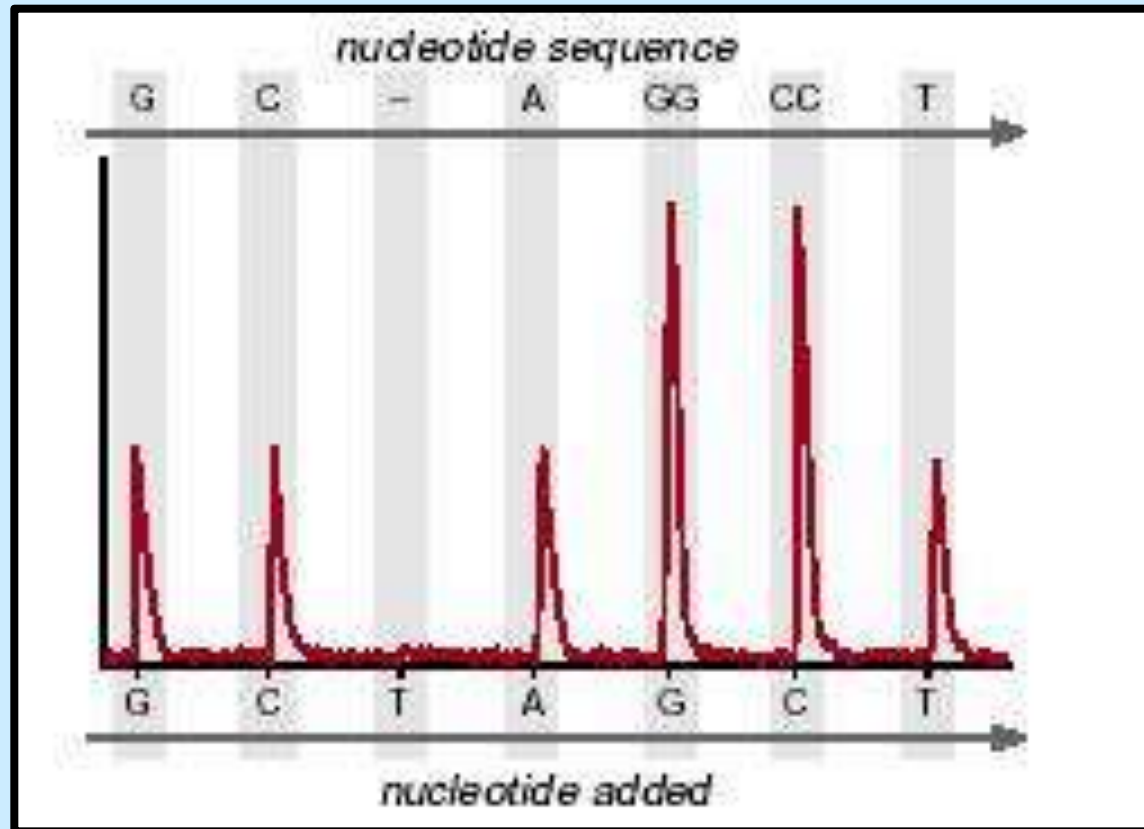
**dXTP → dXMP**

**apyráza**

# *Parametry pyrosekvenování*

- žádné značené primery ani značené nukleotidy
- žádná elektroforéza
- Principem je uvolnění pyrofosfátu při syntéze DNA a emise viditelného světla na konci kaskády reakcí
- Množství uvolněného světla je přímo úměrné množství zabudovaných nukleotidů
- Namísto standardního ATP se používá **2'-deoxyadenosin-5'- $\alpha$ -thio-trifosfát**, který není substrátem pro luciferázu
- Rychlost sekvenování = 1 báze za minutu
- Délka stanovené sekvence = 100 nukleotidů

# Pyrosekvenování - záznam



Intenzita signálu odpovídá počtu začleněných nukleotidů



***Prohlédněte si animaci  
o pyrosekvenování na***

**<http://www.youtube.com/watch?v=jylCHBxTKkw&feature=related>**

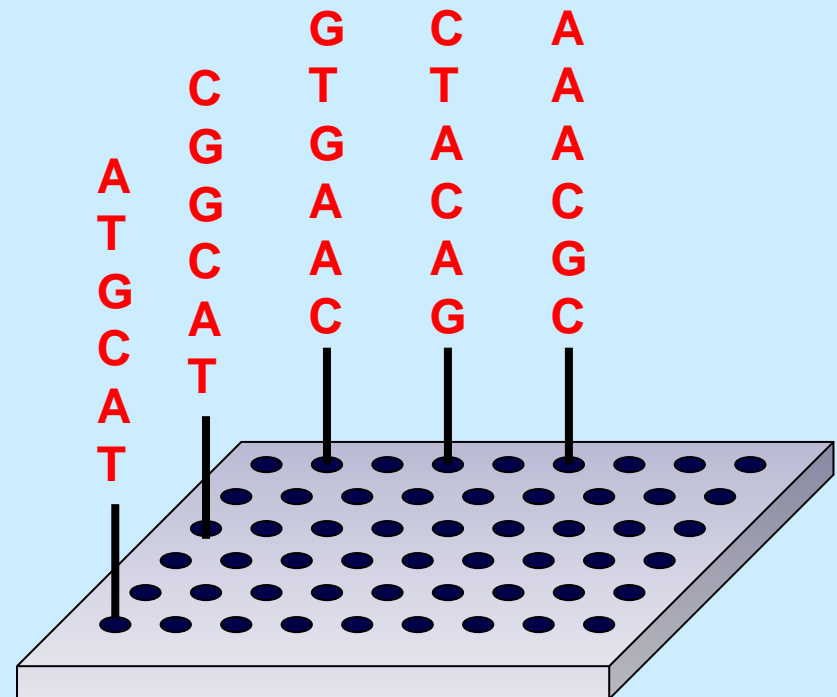


# *Sekvenování pomocí hybridizace*

- **Nepřímá metoda využívající DNA čipů**
- **Výsledek odečten konfokálním mikroskopem**
- **Sekvence odečtena dedukcí**



T  
A  
C  
G  
T  
A



# Otázka



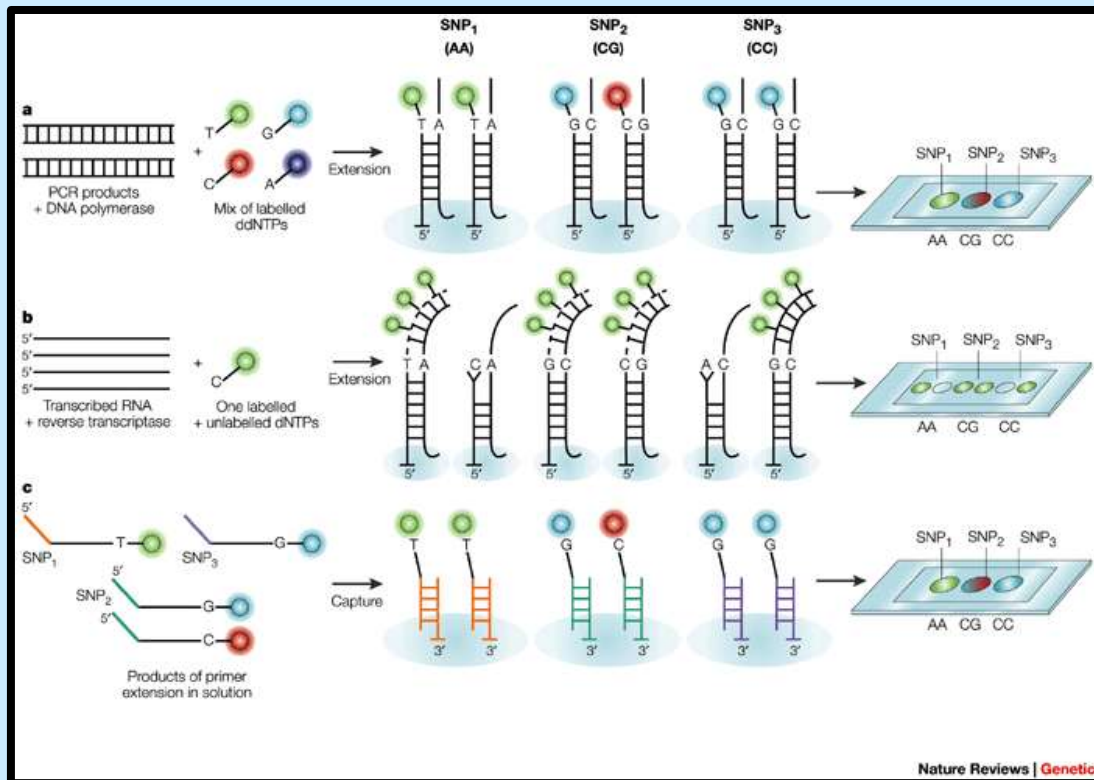
Kdyby byly na čipu naneseny 20 mery, kolik políček by musel obsahovat čip pro sekvenování 1 Mb?

Prý celkem  $1,112 \times 10^6$

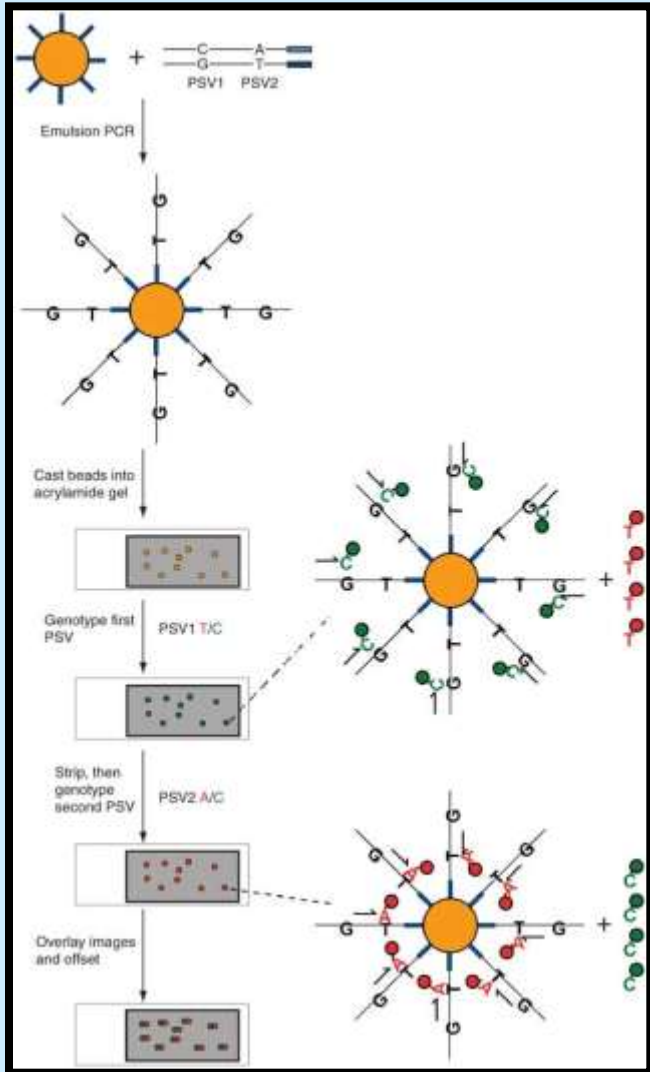


# Minisekvenování

- Metoda využívající technologie DNA čipu
- Využívá se pro stanovení jednoduchých nukleotidových polymorfismů
- Více variant, metoda je automatizovaná, multiplexní



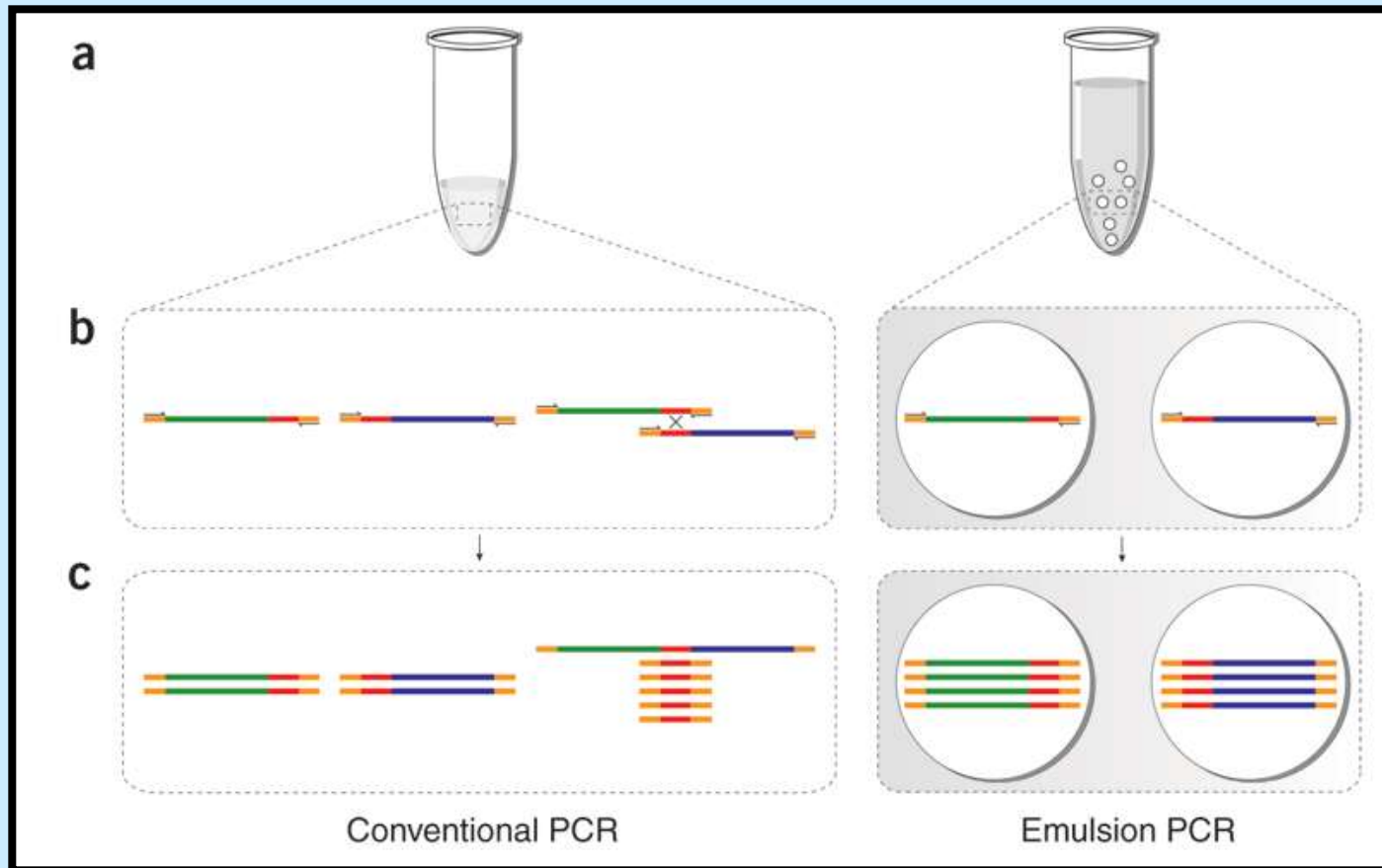
# Emulsní PCR



- Amplifikace probíhá v emulsi (voda-olej)
- DNA je fragmentována na úseky dlouhé 300-800 bp
- Jednotlivé matrice jsou navázány samostatně na povrch jednotlivých kapek obalených primery

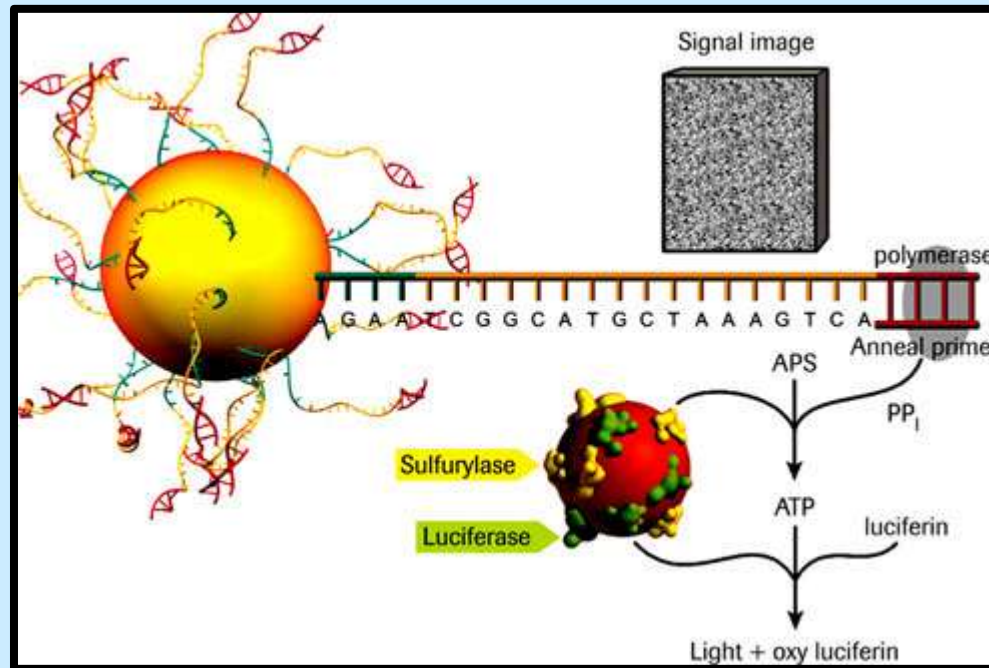
# Emulsní PCR

- Amplikon je výsledkem jedinečné PCR
- Méně nespecifických produktů



# *Komerční aplikace emulsní PCR*

## 454 sekvenční systém firmy Roche



- Produkty PCR jsou sekvenovány pyrosekvenováním

# ***454 sekvenční systém***

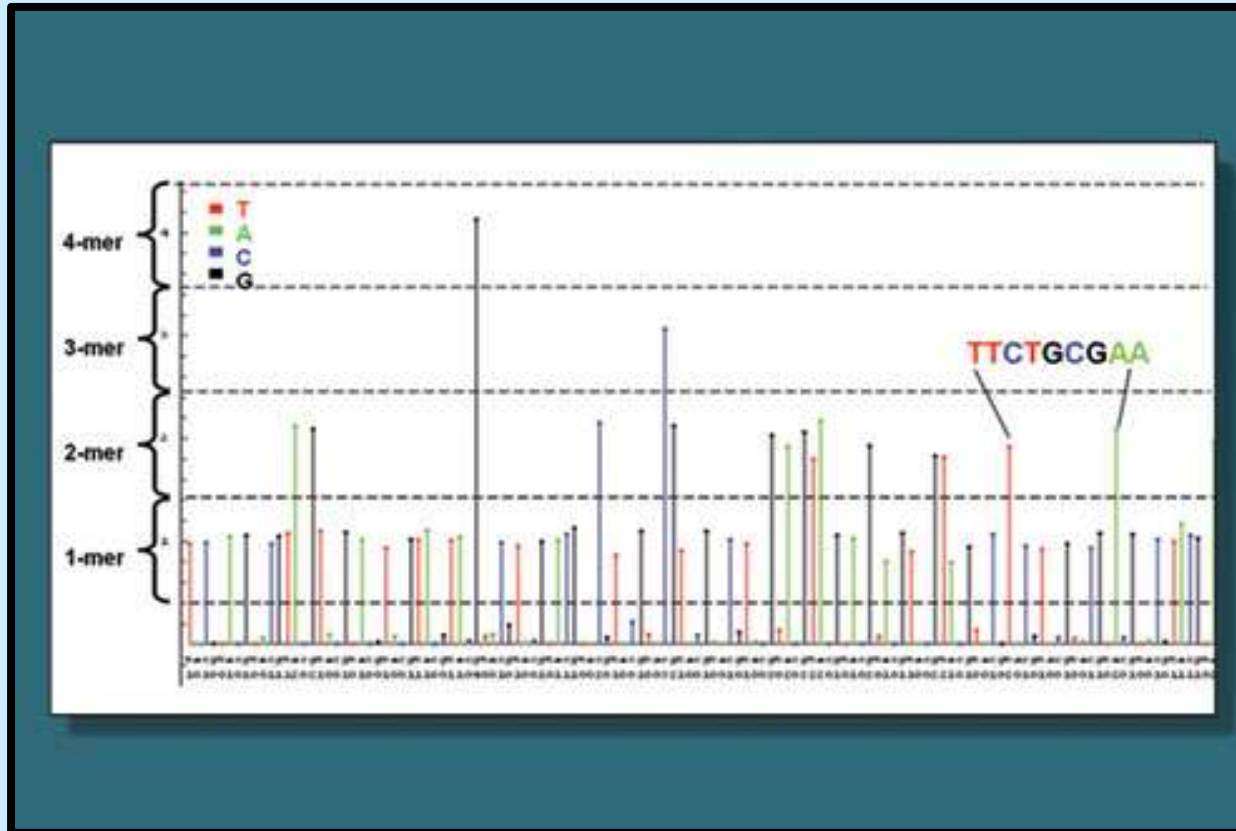
## **Základní kroky**

- **Tvorba jednořetězcových DNA matic**
- **Připojení adaptérů a vazba na pevné částice**
- **Amplifikace DNA matic v emulzi**
- **Vytvoření sekvenčních dat**
- **Analýza sekvencí různými nástroji bioinformatiky**

**Podívejte se na komerční prezentaci na stránkách  
<http://454.com/products/technology.asp>**

# 454 sekvenční systém

## Výsledný záznam





**Další moderní přístupy  
najdete na**

**<http://grf.lishtm.ac.uk/sequencing.htm>**

**SOLiD  
(Applied Biosystems)**

**454/Pyrosequencing  
(Roche)**

**SOLEXA  
(Illumina)**



# ***Sekvenování RNA***

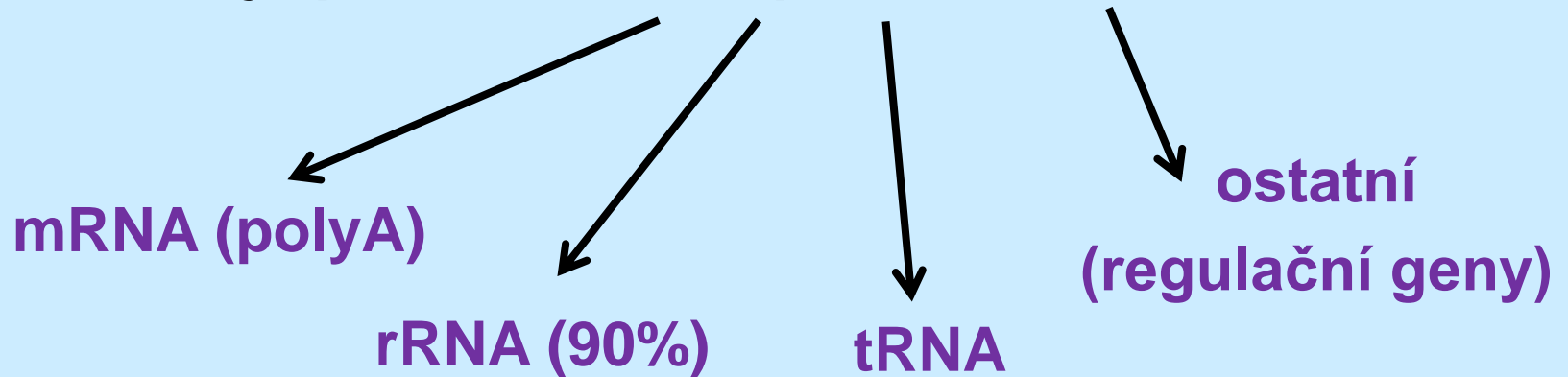
## **Whole Transcriptome Shotgun Sequencing (WTSS)**

- **je možno provést sekvenování cDNA a tím získat informaci o obsahu RNA v daném okamžiku života buňky**

**Morin et al. (2008): Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing". *BioTechniques* 45 (1): 81–94.**

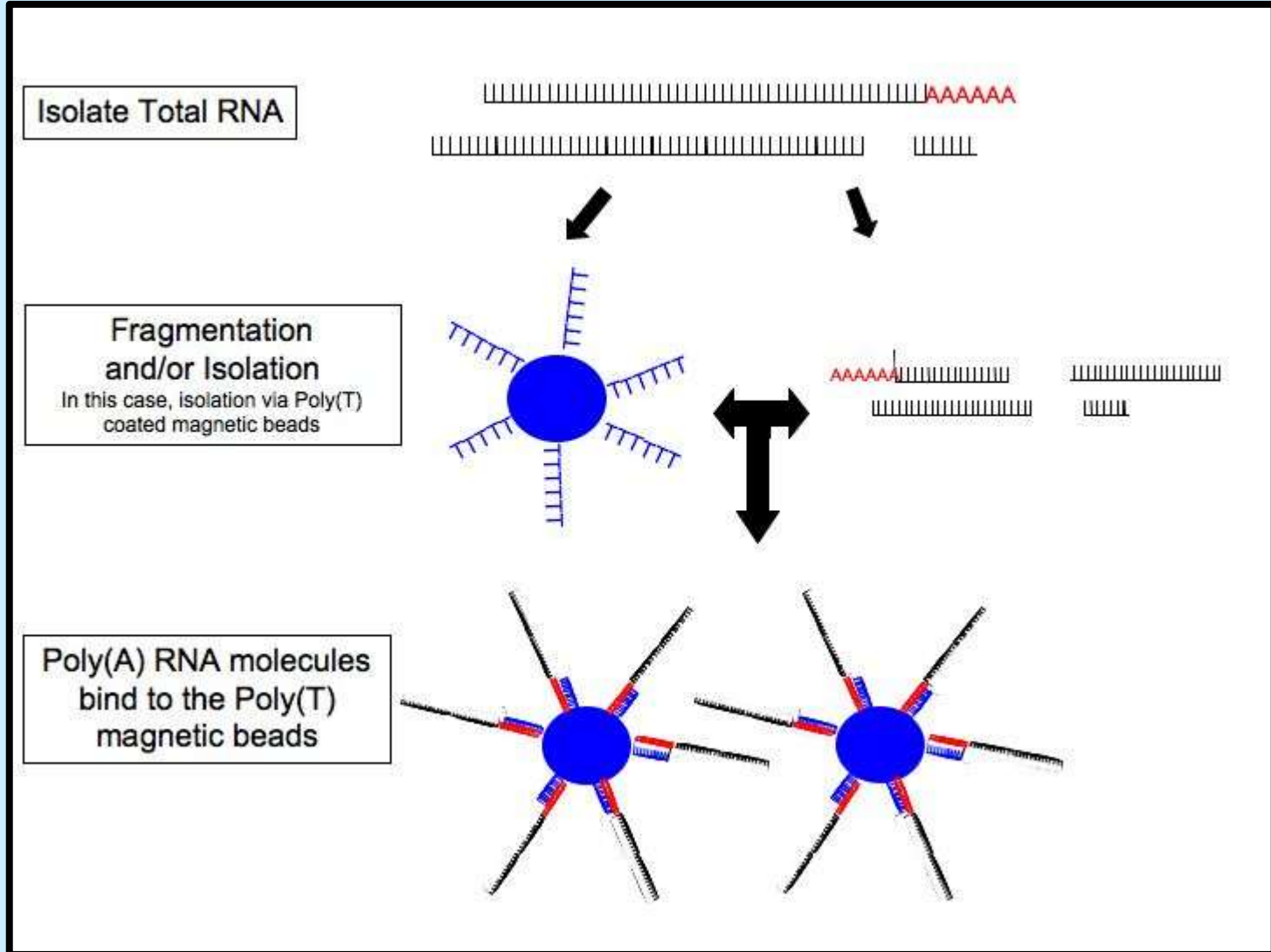
# *Separace různých druhů RNA*

... je prvním krokem při sekvenování RNA



... k separaci se používají nejčastěji magnetické kuličky s poly(T)

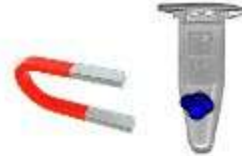
# Separate RNA



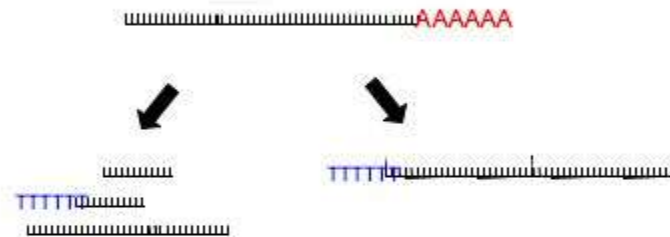


# Zpětná transkripce

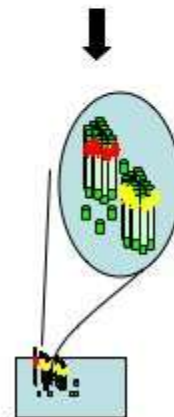
Magnetically isolate  
and wash beads



Fragment and/or Reverse Transcribe



Fragmentation (if not done already),  
size selection, and sequence



Illumina Solexa, Roche 454, or ABI SOLID  
Graphic shown here is Illumina

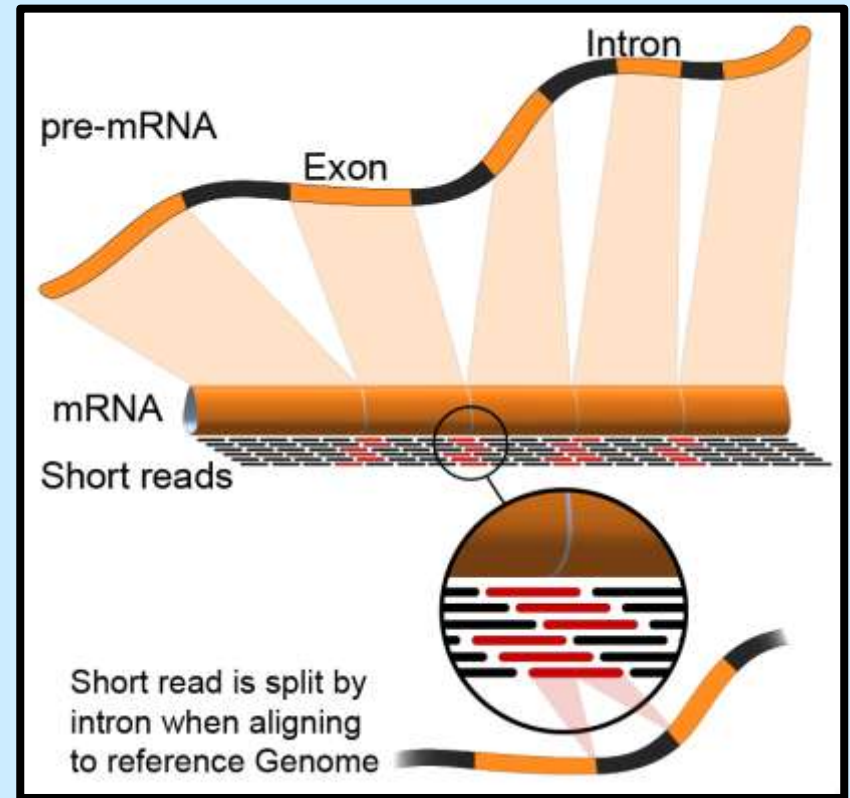
# *Na jaké fragmenty nastříhat RNA?*

<b>Délka čtení</b>	<b>454</b>	<b>Illumina</b>	<b>SOLiD</b>
<b>bp</b>	<b>250</b>	<b>100</b>	<b>35</b>
<b>Mb/run</b>	<b>100 Mb</b>	<b>300 Gb</b>	<b>3 Gb</b>
<b>Doba</b>	<b>7 hod</b>	<b>7-14 dnů</b>	<b>5 dnů</b>

**Podrobnosti k moderním sekvenačním metodám v 5. ročníku**

# Co s vzniklými fragmenty?

- Poskládání do souvislých sekvencí
- Porovnání s genetickými databázemi



Více uslyšíte v 5. ročníku (<http://www.ncbi.nlm.nih.gov/>)



# ***Přímé sekvenování RNA***

**V průběhu přípravy cDNA vznikají artefakty**



**Snaha sekvenovat RNA přímo**



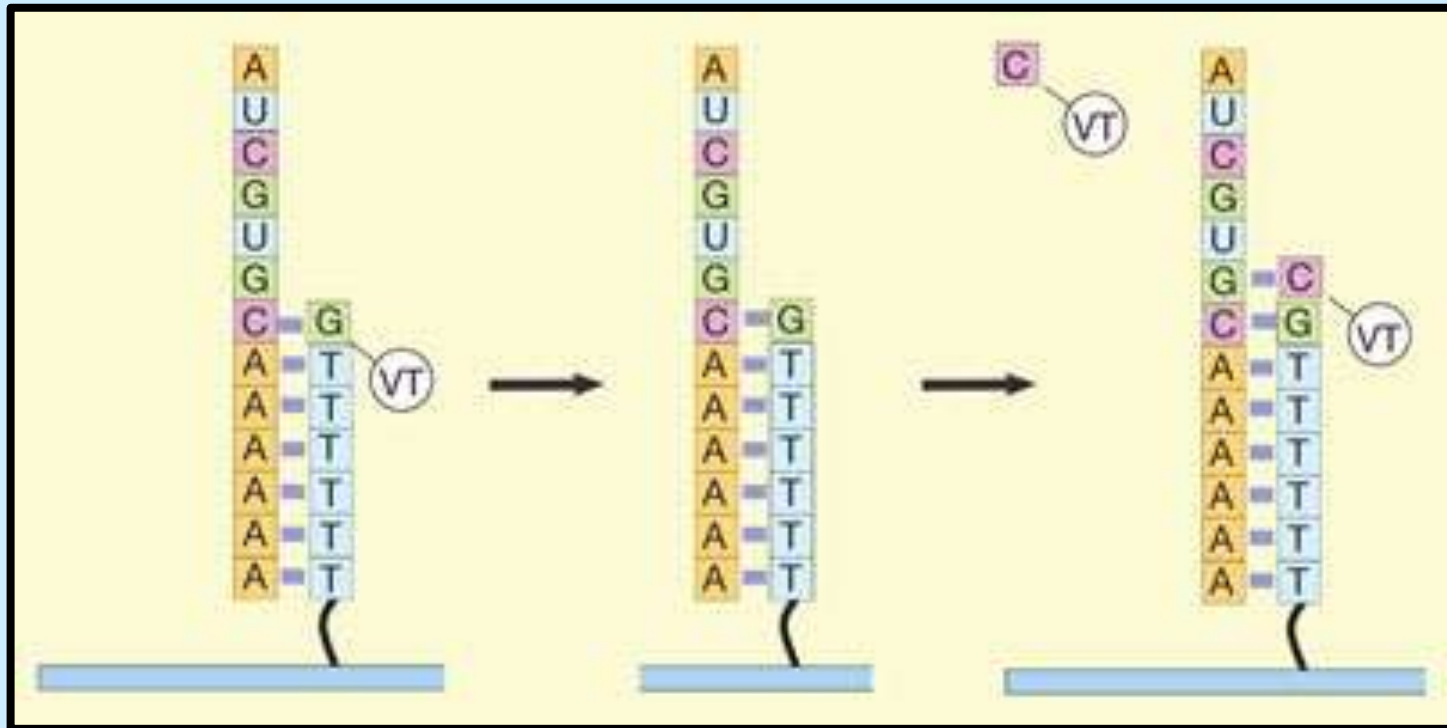
**Direct RNA Sequencing (DRSTM)**



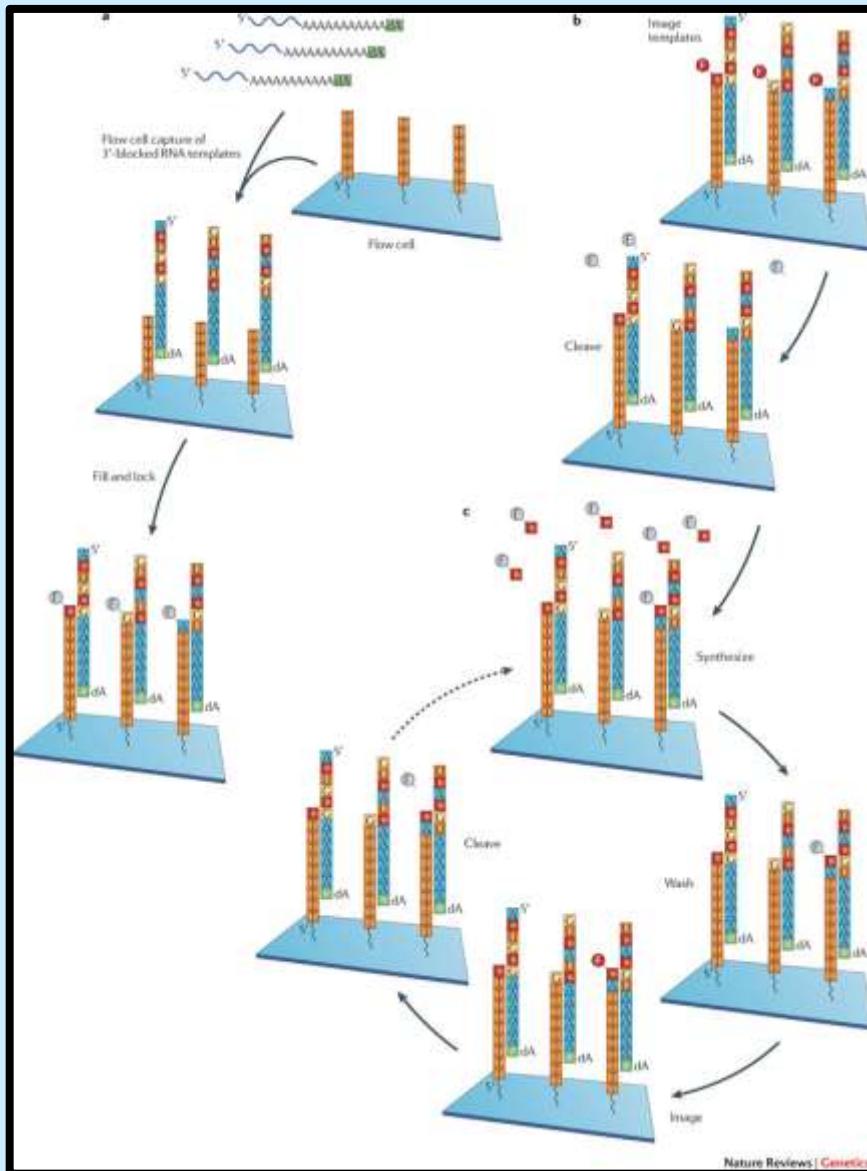
**Helicos Biosciences (<http://www.helicosbio.com/>)**

# *Princip přímého sekvenování RNA*

- Připojení jednotlivé molekuly k ukotvenému oligo(dT)
- Polymerace virtuálního terminátoru



# Průběh přímého sekvenování RNA



- Připojení k oligo(dT)
- Fixace a uzamčení
- Polymerace NTP s VT
- Odmytí zbylých NTP
- Zachycení fluorescence
- Odštěpení VT
- Polymerace dalšího NTP s VT

# ***Využití přímého sekvenování RNA***

Prvně použita k sekvenování mRNA ze *S. cerevisiae*

Oszolak et al. (2009): Direct DNA sequencing. Nature  
8;461(7265):814-8

## **Firma HelicosBiosciences doporučuje**

- **Kvantitativní mapování polyA/Digitální genová exprese**
- **Analýza transkriptomu**
- **Kvantifikace RNA ve formalínových a parafínových preparátech**
- **Malé RNA/nekódující RNA**
- **Imunoprecipitace RNA**
- **Charakterizace RNA na úrovni atomolů**

# Otázka



**Jeden attomol, kolik je to molekul RNA o délce 300 a 500 nukleotidů?**

**1 mol =  $6,023 \times 10^{23}$  molekul**

**$1 \times 10^{-18}$  =  $6,023 \times 10^5$  molekul**



# *Pro náruživé statistiky*



## ***Statistical Design and Analysis of RNA Sequencing Data***

**Auer a Doerge (2010): Statistical Design and Analysis of RNA Sequencing Data. Genetics 185: 405–416**

**Novinky ze sekvenování RNA  
najdete na**

**<http://www.rna-seqblog.com/>**



**Ale je to spíš o vyšších  
eukaryotech**



# ***Komerční aplikace***

## **Illumina (TruSeq RNA Sample Preparation Kits)**

**<http://www.illumina.com/applications/sequencing/rna.ilmn?source=transcriptome>**



## **Invitrogen (Ion Total RNA-Seq Kit)**

**<http://products.invitrogen.com/ivgn/product/4475936>**







***Prohlédněte si animaci  
o Illumina na***

**<http://www.youtube.com/watch?v=l99aKKHcxC4&feature=related>**

# *Otázka z června 2012*

## **Zdroj**

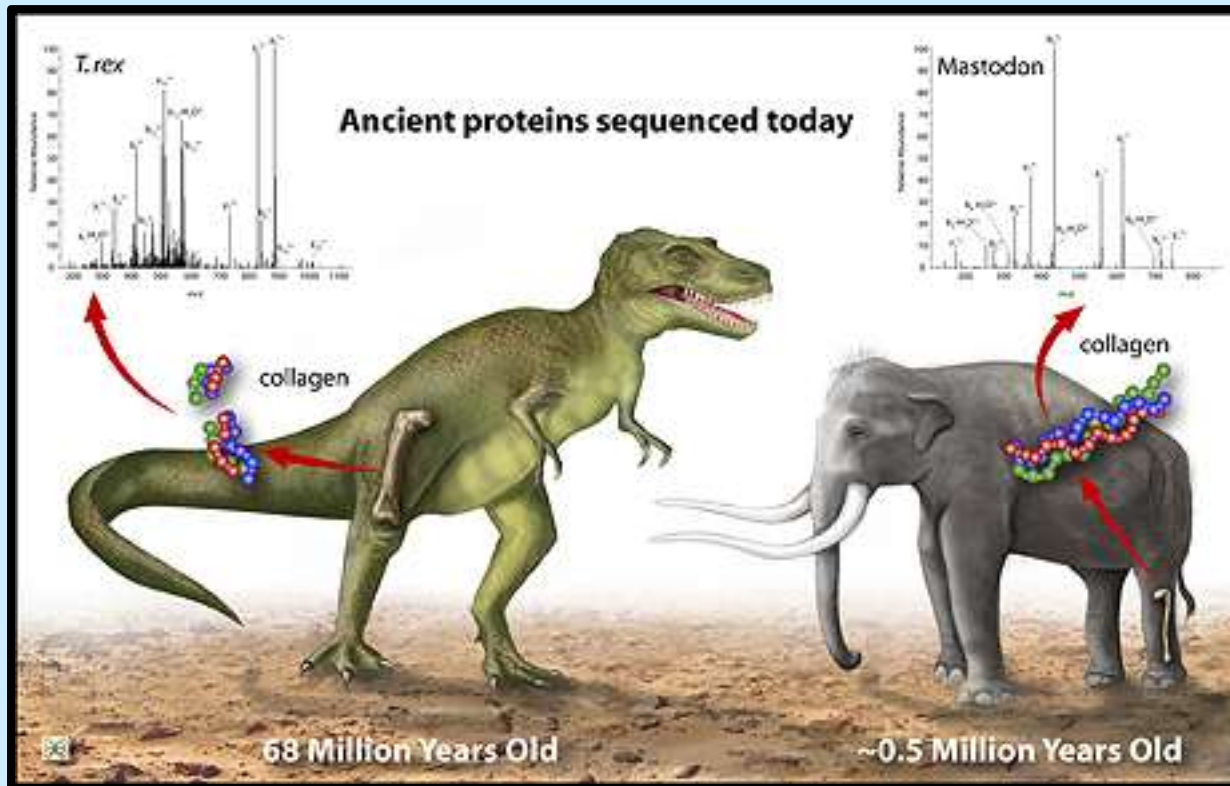
<http://seqanswers.com/forums/showthread.php?t=21016>

Can we separate virus using small RNA sequencing?

---

**We have a plant material which infected virus. Now we want to know the detail information about the virus. How should we do? Can we use small RNA sequencing to separate it? Thank you!**

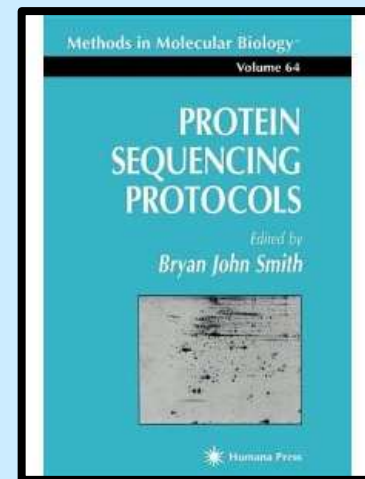
# *Sekvenování proteinů*



# Sekvenování proteinů

... je stanovení sekvence aminokyselin v polypeptidovém řetězci

- První protein, bovinní inzulín byl sekvenován v roce 1953 Frederikem Sangerem
- Sanger obdržel Nobelovu cenu v roce 1958



# ***Strategie sekvenování***

- **Zjistit počet polypeptidových řetězců (podjednotek)**
- **Určit počet disulfidických můstků (uvnitř řetězce a mezi řetězci)**
- **Stanovit sekvenci aminokyselin každého z řetězců**
  
- **Je-li podjednotka příliš dlouhá, fragmentovat na kratší polypeptidy**
- **Sekvenovat fragmenty metodou podle Edmana**
- **Poskládat sekvence analýzou překryvů**

# ***Analýza koncových skupin***

- **Počet podjednotek lze zjistit podle počtu N a C konců**
- **Analýza N konců**
  - Dansyl chlorid
  - Fenylisothiokyanát/ Edmanovo reagens
  - Aminopeptidáza
- **Analýza C konců**
  - Karboxypeptidáza

**Převěd'te úsek DNA do sekvence aminokyselin,  
vyznačte C a N konec**



kódující DNA-řetězec:

5' - ATG AAA TAC GCT CCC TTA AAA - 3'

antikódující DNA-řetězec:

3' - TAC TTT ATG CGA GGG AAT TTT - 5'

Tabulka genetického kódu:

AAA – Lys

AAU – Asn

ACU – Thr

AUG – Met

CGA – Arg

CGU – Arg

CCC – Pro

GCA - Ala

GCU - Ala

GGG - Gly

UAC - Tyr

UGA – Term

UUA - Leu

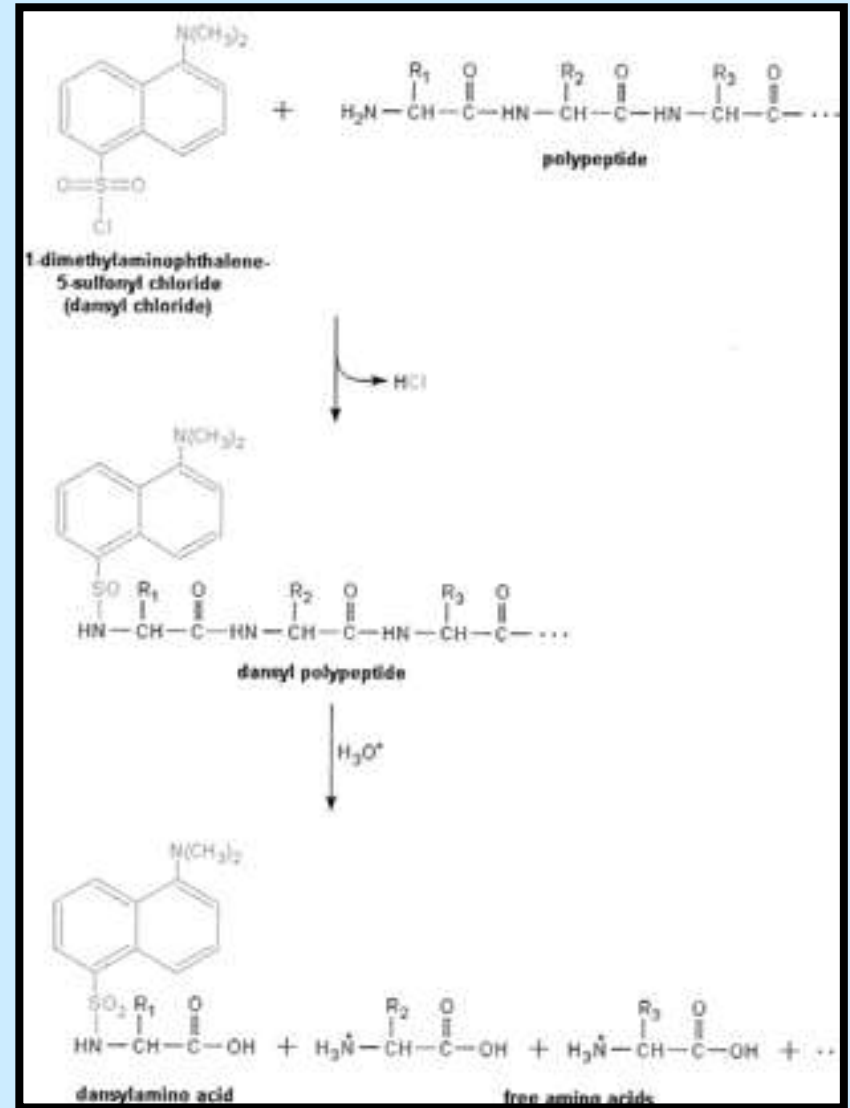
UUU - Phe

**N** - Met - Lys - Tyr - Ala  
- Pro - Leu - Lys - **C**



# Analýza N konců dansyl chloridem

- Hlavní reagensie: 1-dimetyl aminoftalen-5-sulfonyl chlorid (dansyl chlorid)
- Příprava dansylovaného polypeptidu
- Kyselá hydrolyza – uvolnění všech AA a N-koncové dansylované AA
- Separace aminokyselin
- Detekce fluoreskující dansylované AA
- Porovnání s dansylovanými standardy AA

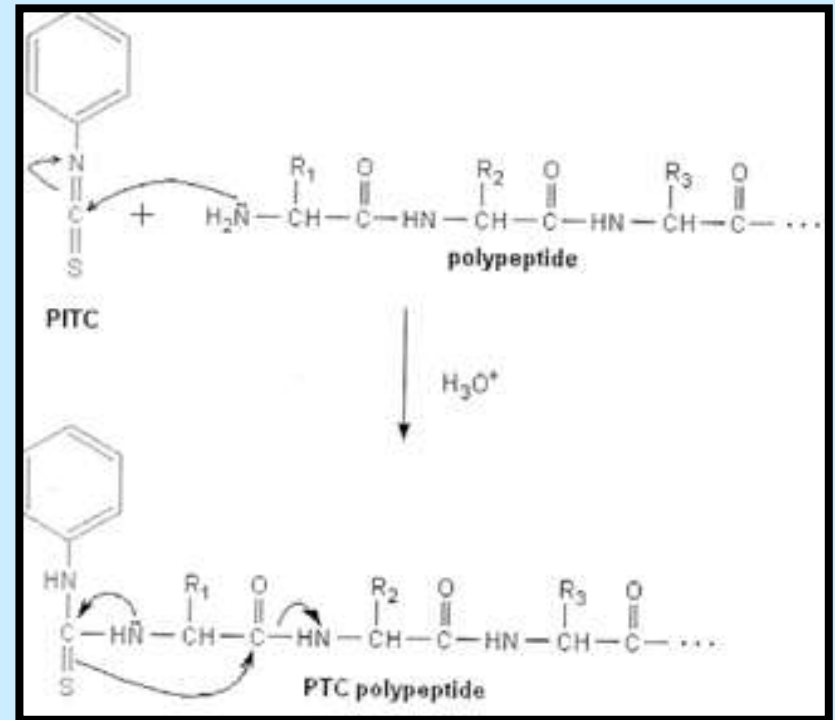




# Analýza N konců podle Edmana I

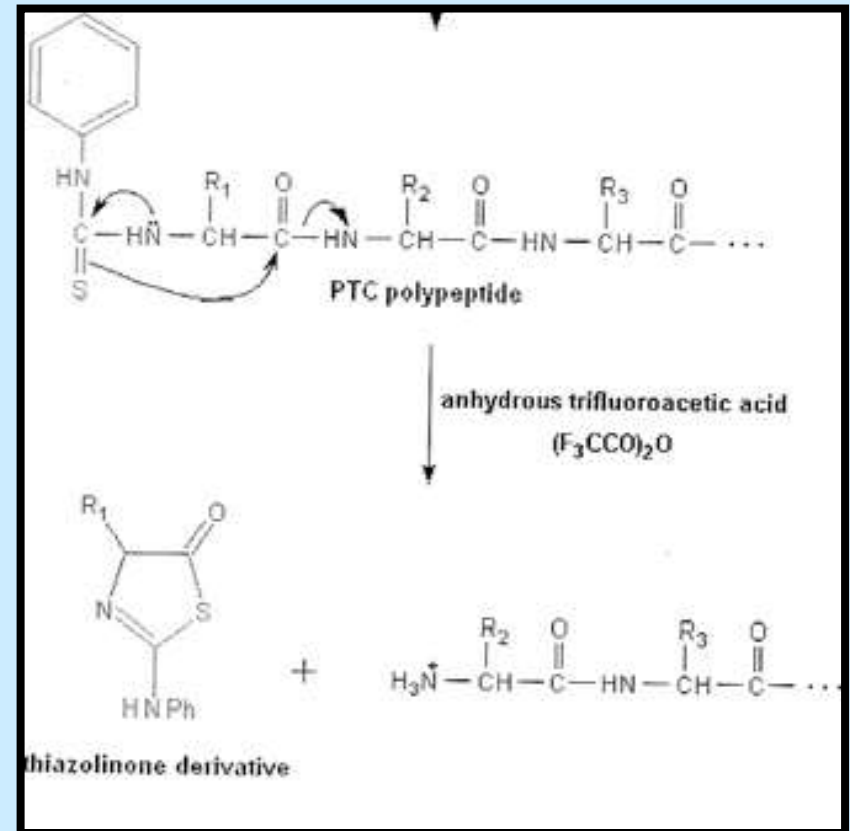
## Metoda degradační

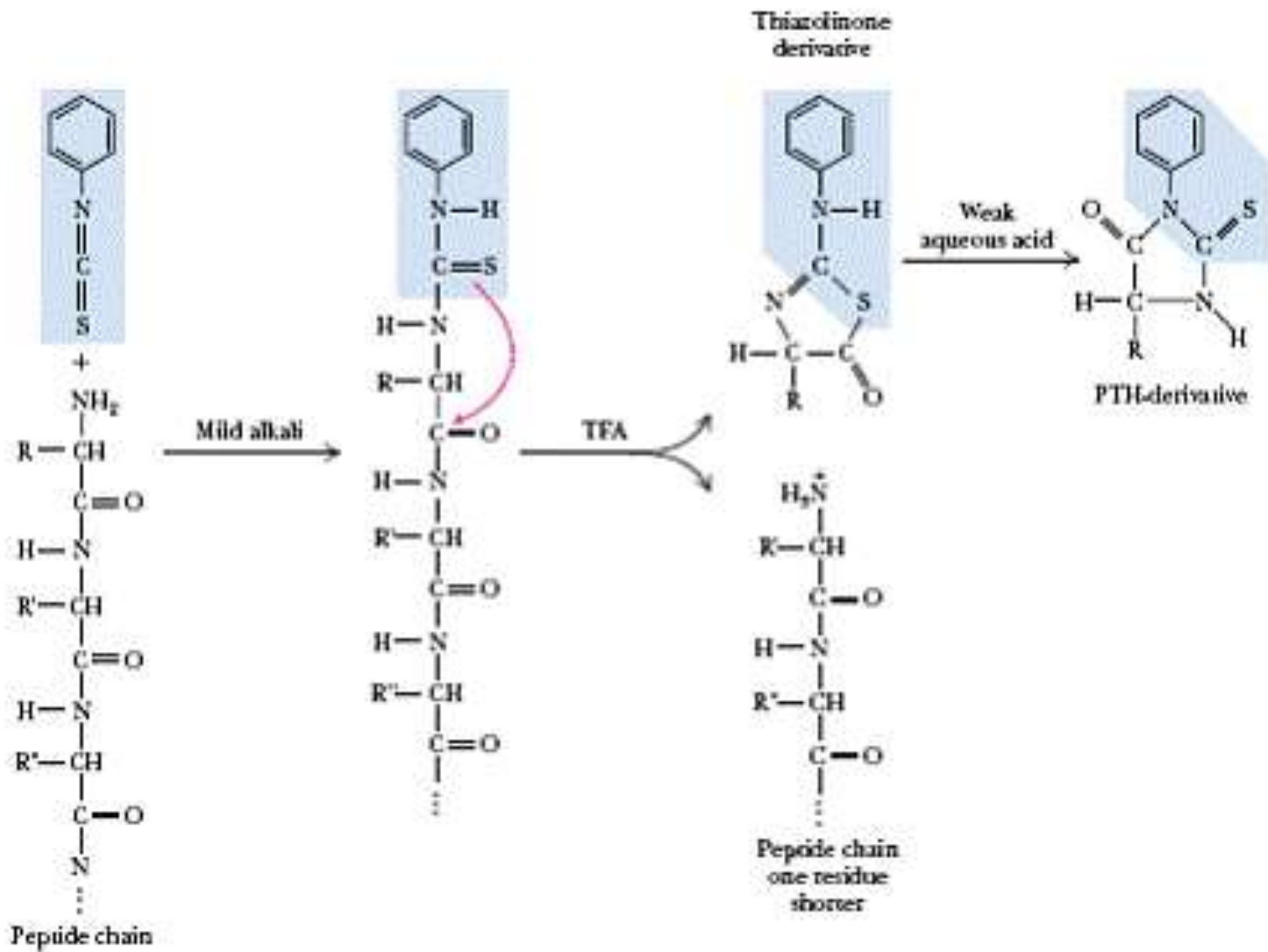
- Nukleofilní atak fenyl izothiokyanátu, Edmanovo reagens, v mírném alkalickém prostředí (N-metylpiperidin/voda/metanol)
- Tvorba fenylcarbamylovo derivátu (PTC-peptid)



# Analýza N konců podle Edmana II

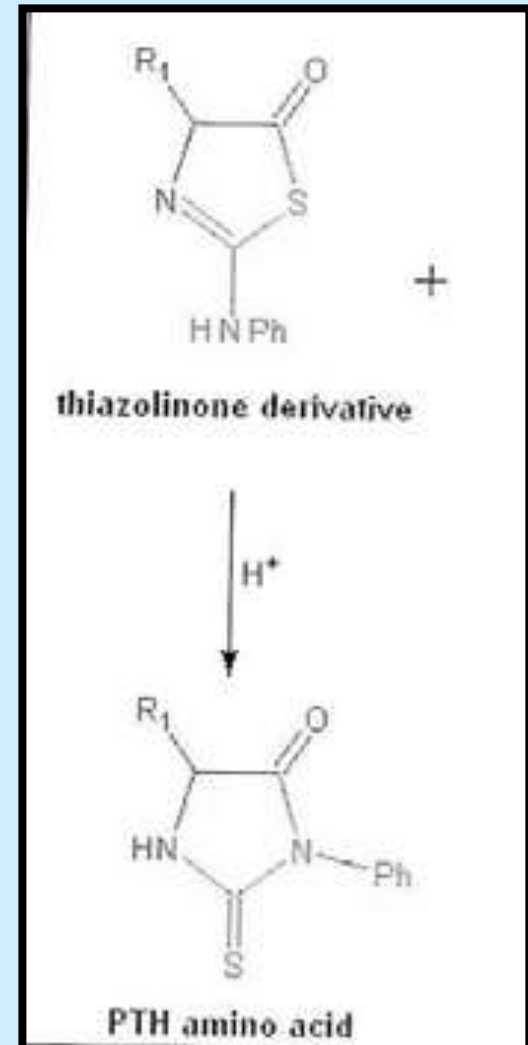
- Kyselina trifluoro octová (TFA) štěpí koncovou aminokyselinu – vzniká thiozolinový derivát a zbylý nemodifikovaný peptid
- Thiozolinový derivát je extrahován organickým rozpouštědlem (např. N-butyl chloridem)
- Zbylý nemodifikovaný peptid nese volnou koncovou aminokyselinu





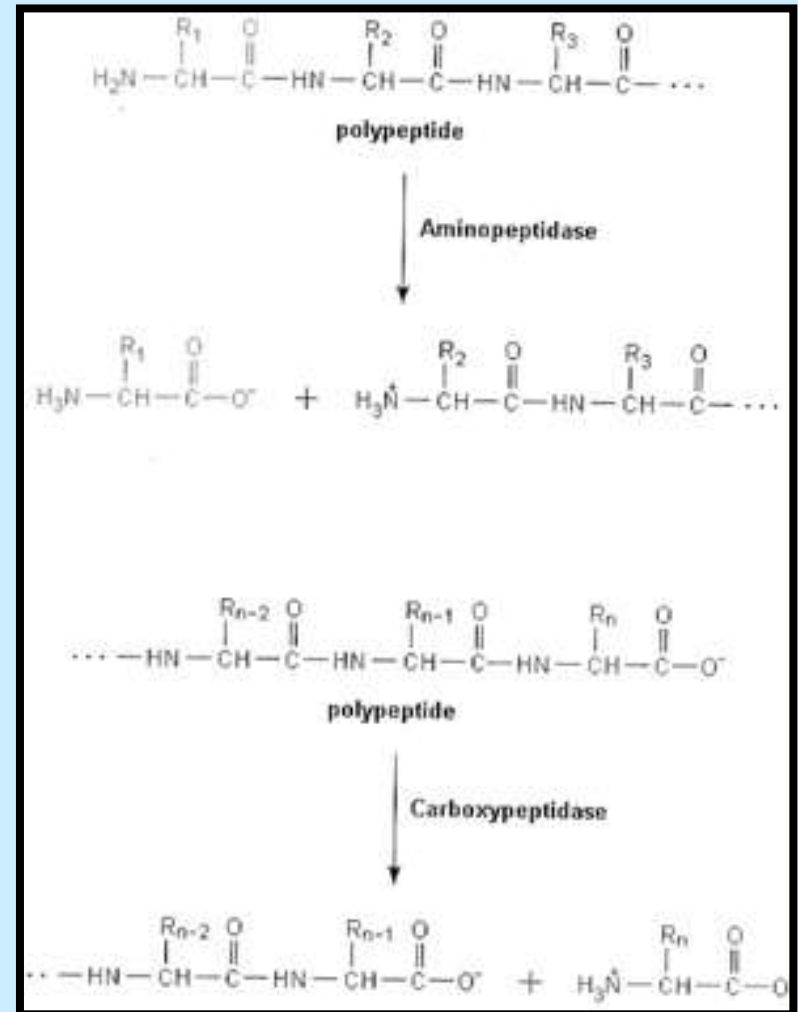
# Analýza N konců podle Edmana III

- Thiozolinový derivát extrahovaný v organickém rozpouštědle je opracován kyselinou (25% TFA) – vzniká derivát fenylthiohydantoinu (PTH)
- PTH je detekován na základě absorpce UV při 296 nm
- PTH AA je separována chromatograficky nebo elfo
- AA je detekována podle retenčního času nebo hmotnosti
- **Sekvenci lze opakovat 40-60x**



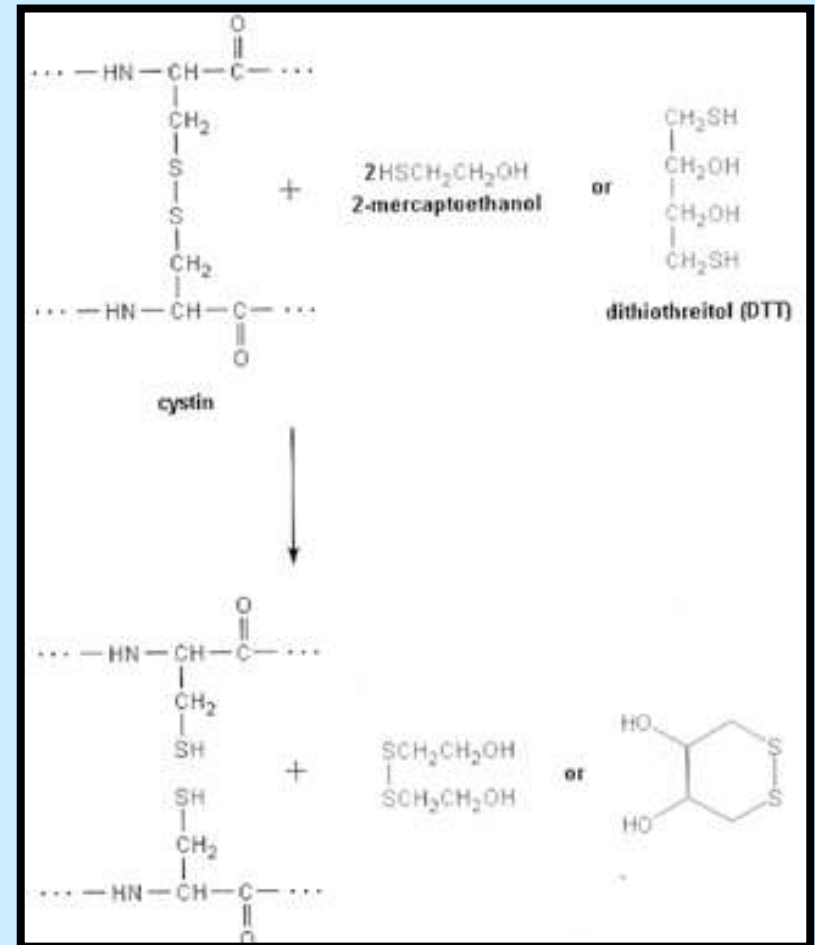
# Analýza N a C konců exopeptidázami

- Exopeptidázy štěpí AA od konců řetězce
- Aminopeptidázy od N-konce
- Karboxypeptidázy od C-konce
- Oba enzymy jsou vysoce specifické, ale pracují pomalu a některé AA jimi nelze odštěpit



# Štěpení disulfidických můstků

- Redukce na thioly prostřednictvím dithiothreitolu nebo 2-merkaptoetanolu
- Thioly jsou opracovány alkylačními činidly (např. kyselinou jodoctovou), aby se zabránilo reoxidaci během následujících kroků



**Účinně sekvenovat lze do 50  
aminokyselin, pak se reakce  
zahltí zplodinami**

**Pracujte s fragmenty**



# *Jak získat fragmenty?*

## Trypsin

- Štěpí pozitivně nabitě AA (Arg nebo Lys), jestliže další AA není prolin
- Štěpí od C-konce

## Endopeptidázy

- Pepsin; štěpí u N-konce Phe, Tyr, Trp není-li předchozí AA prolin
- Chymotrypsin: štěpí u C-konce Phe, Trp, Tyr jestliže další AA není prolin
- Endopeptidáza GluC: štěpí u C-konce Glu



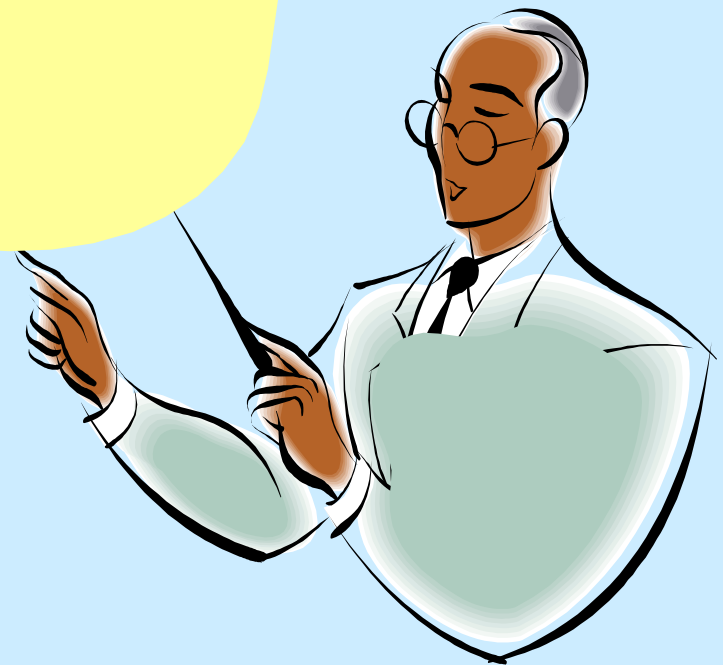
# ***Jak získat fragmenty?***

## **Exopeptidázy**

- **Leucin aminopeptidáza: štěpí N-koncovou AA leucin, neštěpí N-koncový prolin**
- **Aminopeptidáza M: štěpí všechny AA od N-konce**
- **Karboxypeptidáza A: štěpí všechny AA kromě Arg, Lys, and Pro, zvlášt' účinná pro AA s alifatickými a aromatickými postranními řetězci, neštěpí je-li následující AA prolin**
- **Karboxypeptidáza B: štěpí C-koncový Arg a Lys, není-li následující AA prolin**
- **Karboxypeptidáza C: štěpí aminokyseliny od C-konce**

**Existují i chemické metody  
štěpení uvnitř  
polypeptidových řetězců**

**Cyanogen bromid**



# ***Jak rozdělit fragmenty?***

## **Tradiční metody**

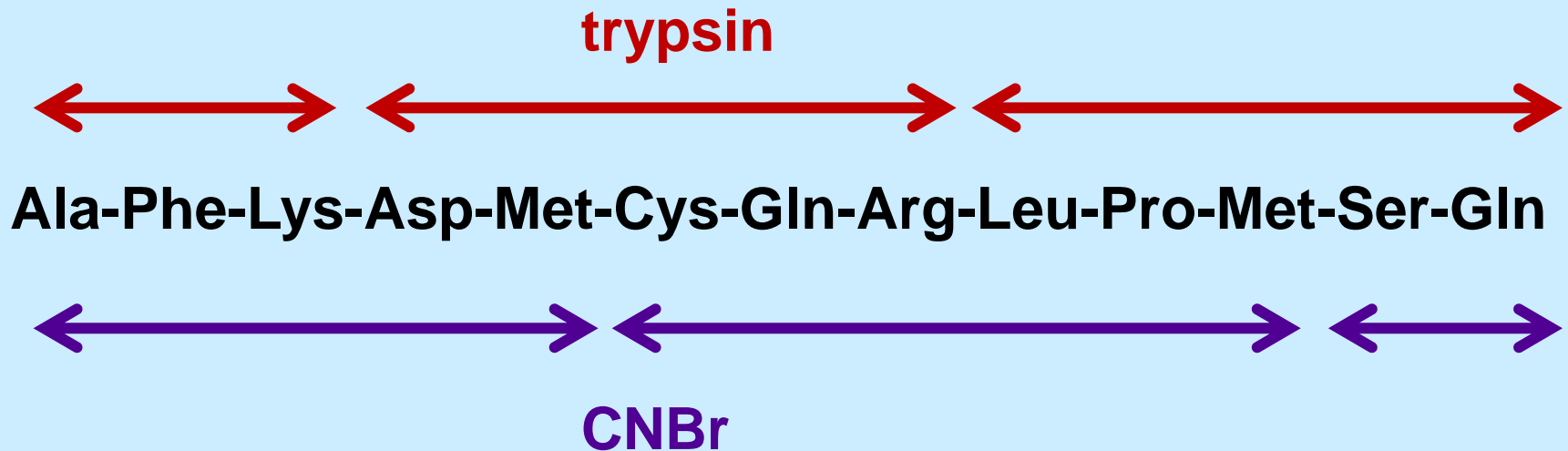
- **Separace podjednotek po štěpení S-S můstků metodou SDS-PAGE nebo HPLC**
- **Zapotřebí hmotnostní standardy a kalibrační křivka**
- **Přibližný počet AA ve fragmentu lze určit z molekulové hmotnosti fragmentu/110**

## **Moderní postupy**

- **MALDI – přesnější a rychlejší**

# *Stanovení sekvence aminokyselin*

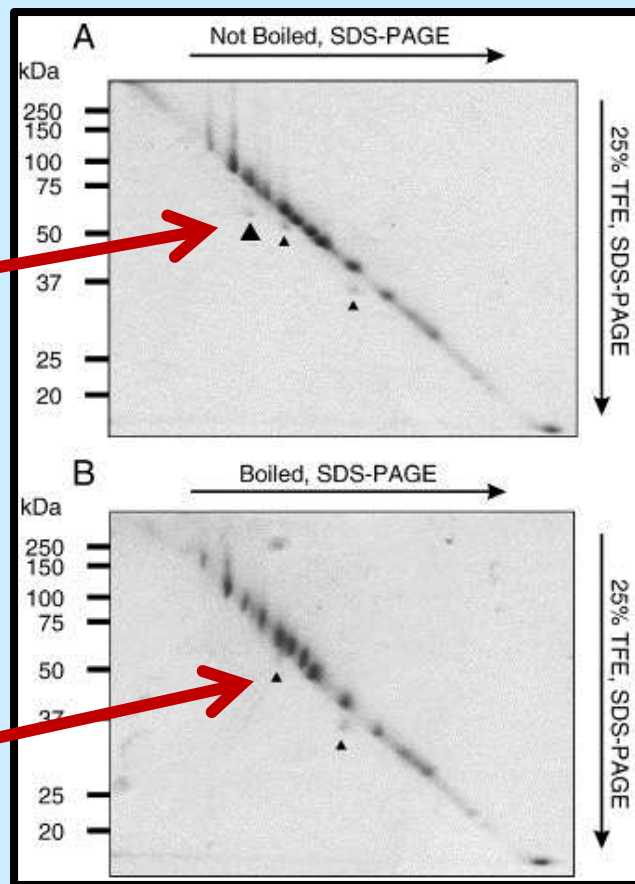
- 1) Získání fragmentů více metodami
- 2) Sekvenování jednotlivých fragmentů
- 3) Poskládání fragmentů s využitím překryvů



# ***Stanovení pozice S-S můstku***

- 1) Fragmentace polypeptidových řetězců**
- 2) 2D gel směsi fragmentů, stejné podmínky v obou rozměrech**
- 3) Po separaci v prvním rozměru opracovat kyselinou permravněčí, která štěpí všechny S-S můstky**
- 4) Separace ve druhém rozměru**
  - Fragmenty bez S-S můstku se rozmístí podél diagonály**
  - Fragmenty s S-S můstky vytvoří spoty mimo diagonálu**
  - Fragmenty s S-S můstky lze z gelu extrahovat a sekvenovat**

# *Příklad analýzy proteinových komplexů*

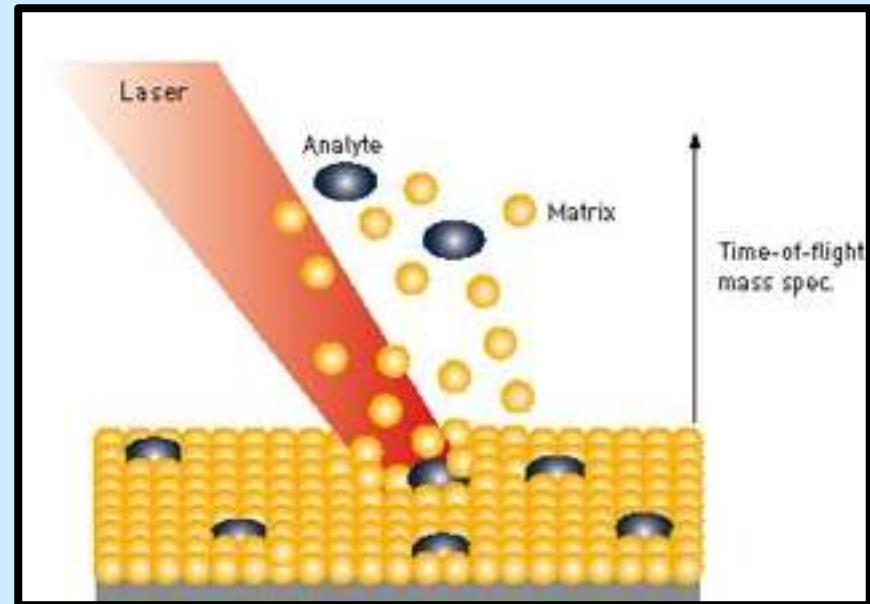


**Gubbens et al. (2008): Protein complexes in bacterial and yeast mitochondrial membranes differ in their sensitivity towards dissociation by SDS. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* 1784 (12), 2012–2018.**

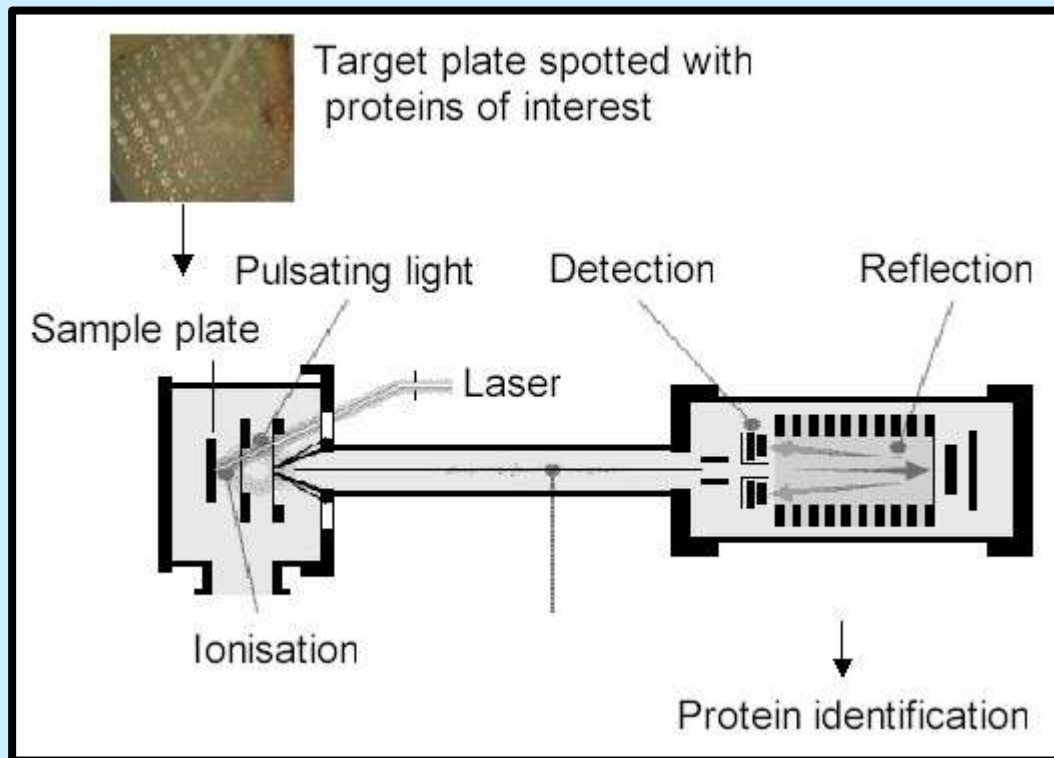
# Sekvenování MALDI-TOF

matrix-assisted laser desorption ionization time-of-flight

- varianta hmotnostní spektrometrie
- peptidy jsou ionizovány a stanoví se poměr hmoty k náboji na základě doby letu (time-of-flight) k detektoru
- vypočte se  $M/z$  a ta je specifická pro každou aminokyselinu

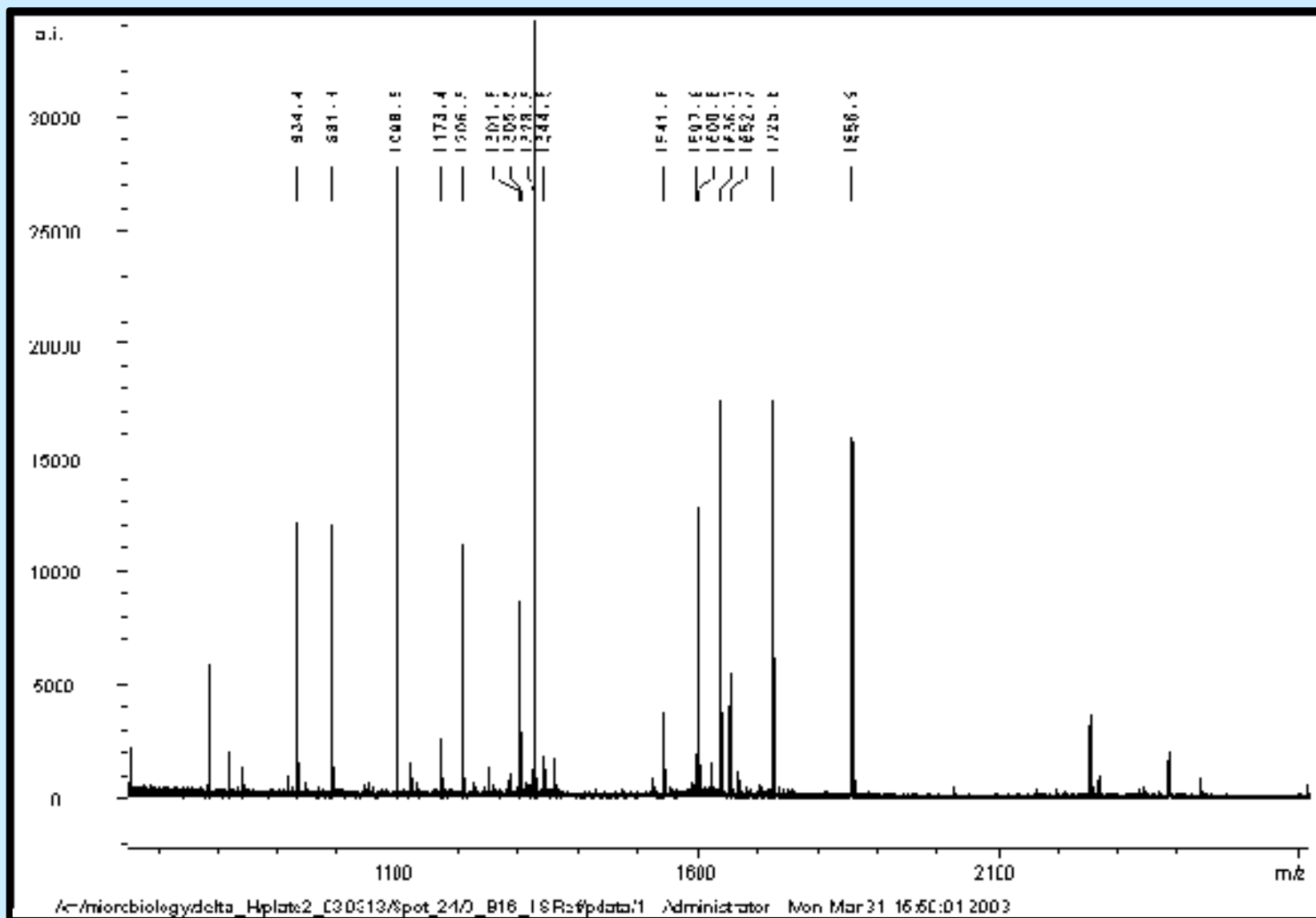


# Schéma zařízení





# Záznam z MALDI-TOF



# *Shotgun protein sequencing*

- Tento koncept vycházející z analogie pro sekvenování DNA se objevil v roce 2007
- Byla zpracována data získaná ze směsi proteinů sekvenovaných hmotnostní spektrometrií
- Analýza podobně jako v případě práce s DNA

**Bandeira et al. (2007): Shotgun Protein Sequencing. *Molecular & Cellular Proteomics* 6:1123–1134.**

# ***Kolik materiálu potřebujeme na analýzu?***

- **limit je kolem 2 pmol, tj. asi  $10^{10}$  molekul**
- **protein by měl být co nejčistší (purifikace HPLC)**

# ***Shrnutí***

- 1) Pojem sekvenování**
- 2) Maxamovo-Gilbertovo sekvenování**
- 3) Sangerova metoda s využitím fluoroforů**
- 4) Pyrosekvenování**
- 5) Nové přístupy k sekvenování**
- 6) Sekvenování RNA**
- 7) Sekvenování proteinů**