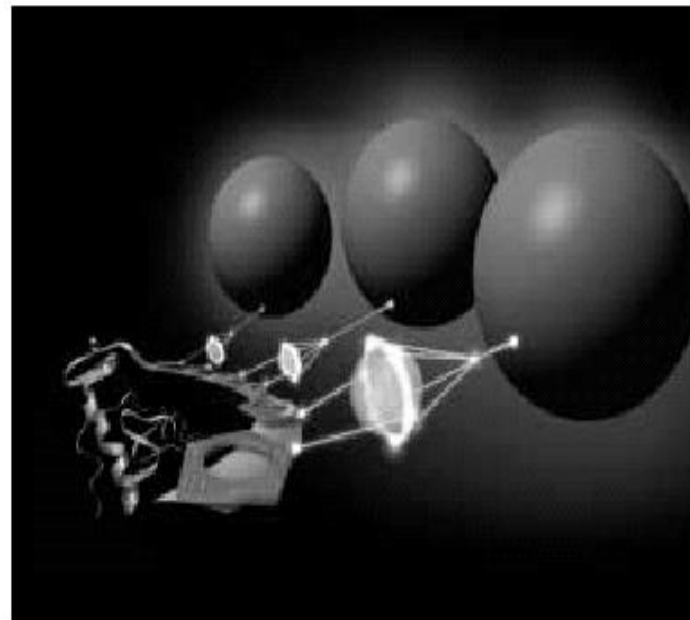
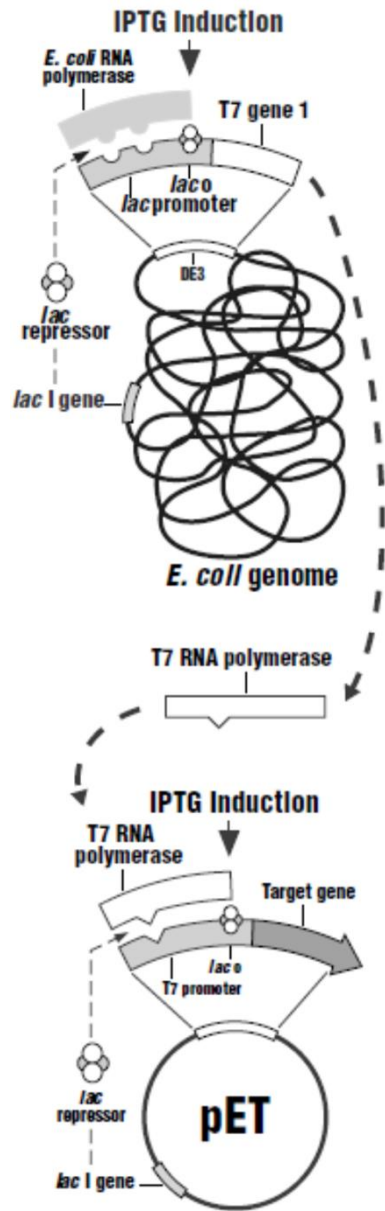


# Expresa a purifikace rekombinantních protein



Radka Dopitová



# Rekombinantní proteiny

**Rekombinantní DNA** je arteficiální DNA sekvence, která vznikla novou kombinací různých DNA sekvencí.

**Rekombinantní proteiny** jsou proteiny získané vnesením rekombinantní DNA do heterologního hostitele (např. mikroorganismus, kvasinky), ve kterém dojde k expresi genu.

# Využití rekombinantních protein

Nadprodukce a purifikace rekombinantních protein jsou nezbytným předpokladem pro:

É **Biochemickou funkční charakteristiku proteinu** (určení přesných kinetických parametrů  $K_m$ ,  $k_{cat}$  pro enzymy se substrátem,  $K_i$  pro enzymy s inhibítorem,  $K_d$  pro protein - proteinové interakce i ligand -proteinové interakce)

É **Strukturní analýzu** (NMR, krystalografie)

É **Proteinové inženýrství** (zlepšení vlastností proteinů o aktivita, stabilita)

É **V průmyslovém měřítku jsou produkovány léky, vakcíny a potravinové doplňky.**

**Cíl:** Vysoký výtěžek homogenního proteinu (mg o kg proteinu)

Zachování biologické aktivity

# Pro výrobu rekombinantní proteiny?

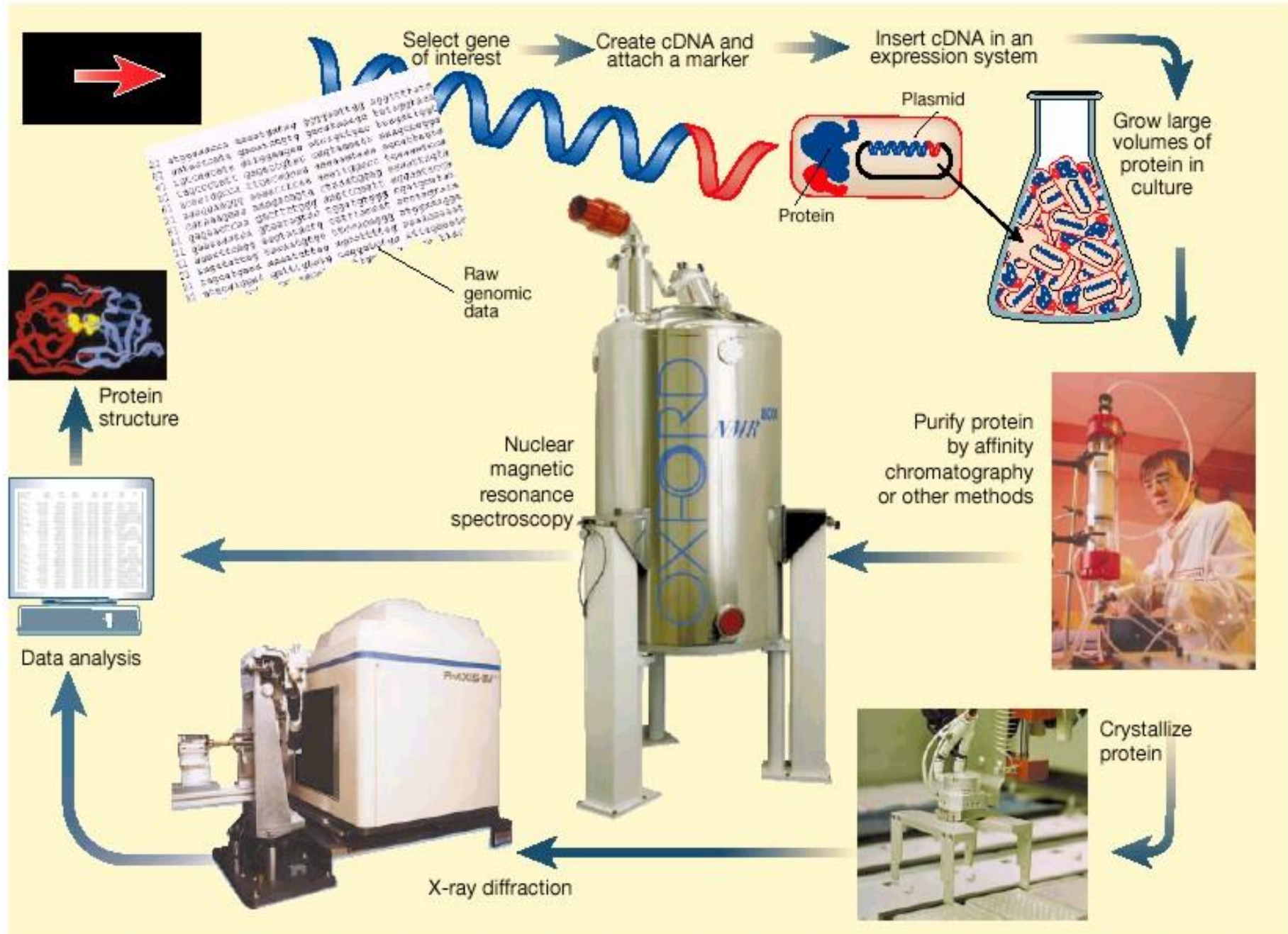
- Přirozený zdroj:**
- É Obtížně se získává (tkáně, orgány).
  - É Obtížně se kultivuje (bakterie, viry, tkáňové kultury).
  - É Limitovaná exprese
  - É Často obtížná purifikace proteinu

TABLE 1.2. Examples of low-abundance proteins and peptides isolated from natural biological sources

Protein	Source	Yield (μg)	Reference
Multipotential colony-stimulating factor	pokeweed mitogen-stimulated mouse spleen-cell-conditioned medium (10 liters)	1	Cutler et al. (1985)
Human A33 antigen	human colon cancer cell lines (10 <sup>10</sup> cells)	2.5	Catimel et al. (1996)
Platelet-derived growth factor (PDGF)	human serum (200 liters)	180	Heldin et al. (1981)
Granulocyte-colony-stimulating growth factor (G-CSF)	mouse lung-conditioned medium (3 liters)	40	Nicola et al. (1983)
Granulocyte-macrophage colony-stimulating growth factor (GM-CSF)	mouse lung-conditioned medium (3 liters)	12	Burgess et al. (1986)
Coelenterate morphogen	sea anemone (200 kg)	20	Schaller and Bodenmuller (1981)
Peptide YY (PYY)	porcine intestine (4000 kg)	600	Tatemoto (1982)
Tumor necrosis factor (TNF)	HL60 tissue culture medium (18 liters)	20	Wang and Creasy (1985)
Murine transferrin receptor	NS-1 myeloma cells (10 <sup>10</sup> cells)	20	van Driel et al. (1984)
Fibroblast growth factor (FGF)	bovine brain (4 kg)	33	Gospodarowicz et al. (1984)
Transforming growth factor-β (TGF-β)	human placenta (8.8 kg)	47	Frolik et al. (1983)
Human interferon	human leukocyte-conditioned medium (10 liters)	21	Rubinstein et al. (1979)
Muscarinic acetylcholine receptor	porcine cerebrum (600 g)	6	Haga and Haga (1985)
β <sub>2</sub> -adrenergic receptor	rat liver (400 g)	2	Graziano et al. (1985)

Adapted, with permission, from Simpson and Nice (1989).

# Technologie rekombinantních proteinů



# Hostitelský organismus pro expresi rekombinantních protein

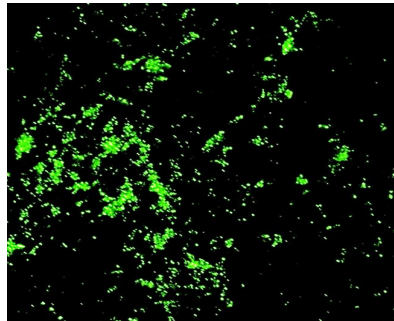
- É Bakterie (nízké náklady, vysoká rychlost růstu, vysoká úroveň exprese)
- É Kvasinky (nízké náklady, vysoká rychlost růstu, nízká aťi vysoká úroveň exprese)
- É Hmyzí buňky (vysoké náklady, nízká rychlost růstu, nízká aťi vysoká úroveň exprese)
- É Savčí buňky (vysoké náklady, nízká rychlost růstu, většinou nízká úroveň exprese)

Expresní systém	Posttranslační modifikace					
	N-glykosylace	O-glykosylace	Fosforylace	Acetylace	Acylace	γ-karboxylace
<i>E. coli</i>	chybí	-	-	-	-	-
Kvasinky	vysoce manosilované glykany	+	+	+	+	-
Hmyzí buňky	jednoduchá, bez sializace	+	+	+	+	-
Savčí buňky	komplexní	+	+	+	+	+

- É In vitro transkripce a translace

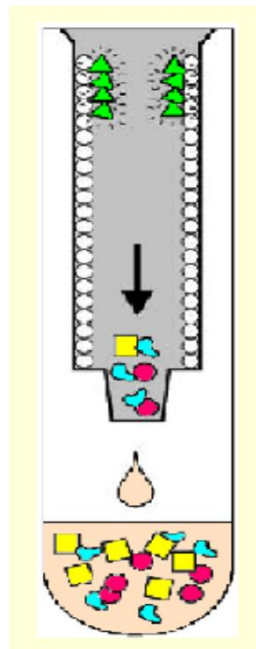
# Obsah přednášky

## 1. část: Expresce rekombinantních proteinů v *E.coli*



Expresce proteinů fúzovaných s GFP v *E. coli*

## 2. část: Purifikace rekombinantních proteinů



Purifikace proteinů fúzovaných s GFP pomocí hydrofóbní matrice

# Produkce heterologních proteinů v *E. Coli*

## VÝHODY :

ÉVysoká produkce rekombinantních proteinů

ÉDobře prostudovaný genom a proteom usnadní genových manipulací

ÉDesign vektorů usnadňuje klonování a expresi cizích genů

ÉRychlý růst v poměrně levném médiu

ÉPřizpůsobivost systému



# Produkce heterologních proteinů v *E. Coli*

## NEVÝHODY:

É Potřeba cDNA zkoumaného proteinu

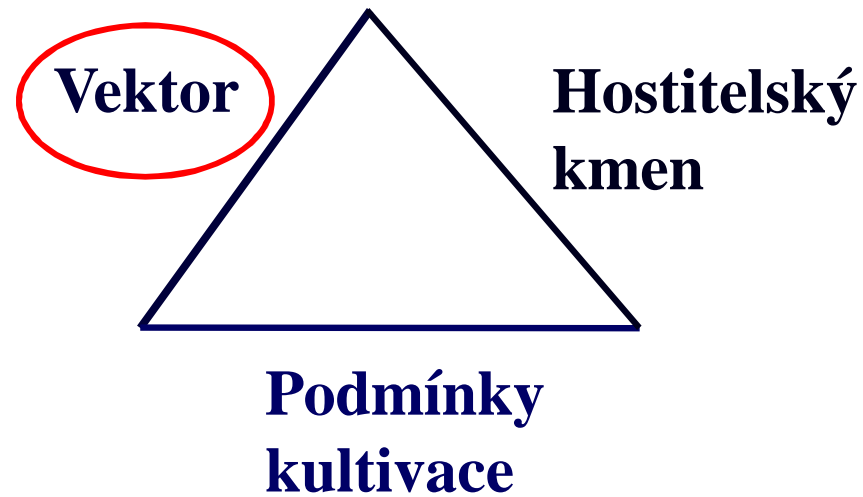
É Absence eukaryotických posttranslačních systémů (posttranslační modifikace)

É Tvorba nerozpustných inkluzních tělísek

É Omezená schopnost tvorby disulfidických vazeb

É Chybí sekreční mechanismus pro účinné uvolnění proteinu do kultivačního média

# Expresní systém pro produkci rekombinantních protein v *E.coli*



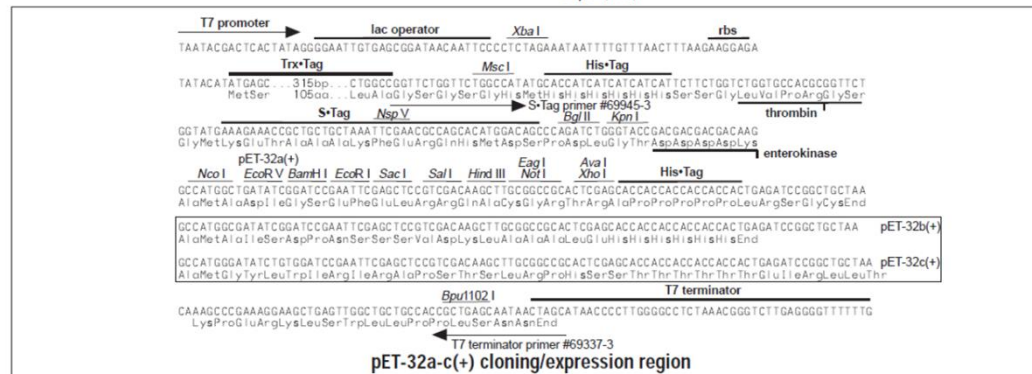
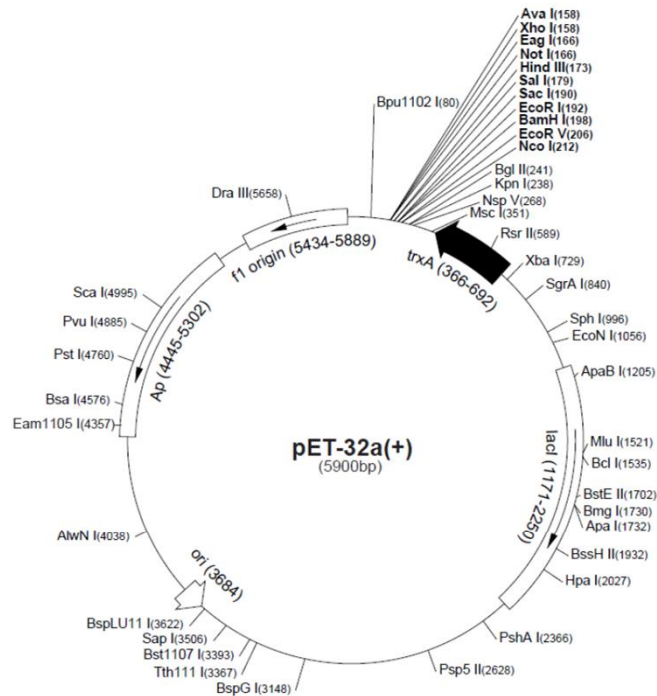
# Expresní vektor

= klonovací vektor, který obsahuje nezbytné regulační sekvence k tomu, aby podporoval expresi inzertních cizích genů.

pET-32a(+) sequence landmarks

T7 promoter	764-780
T7 transcription start	763
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites	
( <i>Nco</i> I - <i>Xho</i> I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
<i>lac</i> I coding sequence	1171-2250
pBR322 origin	3684
<i>bla</i> coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*HI at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*HI at 198 except for *Eco*R V, which cuts at 209.

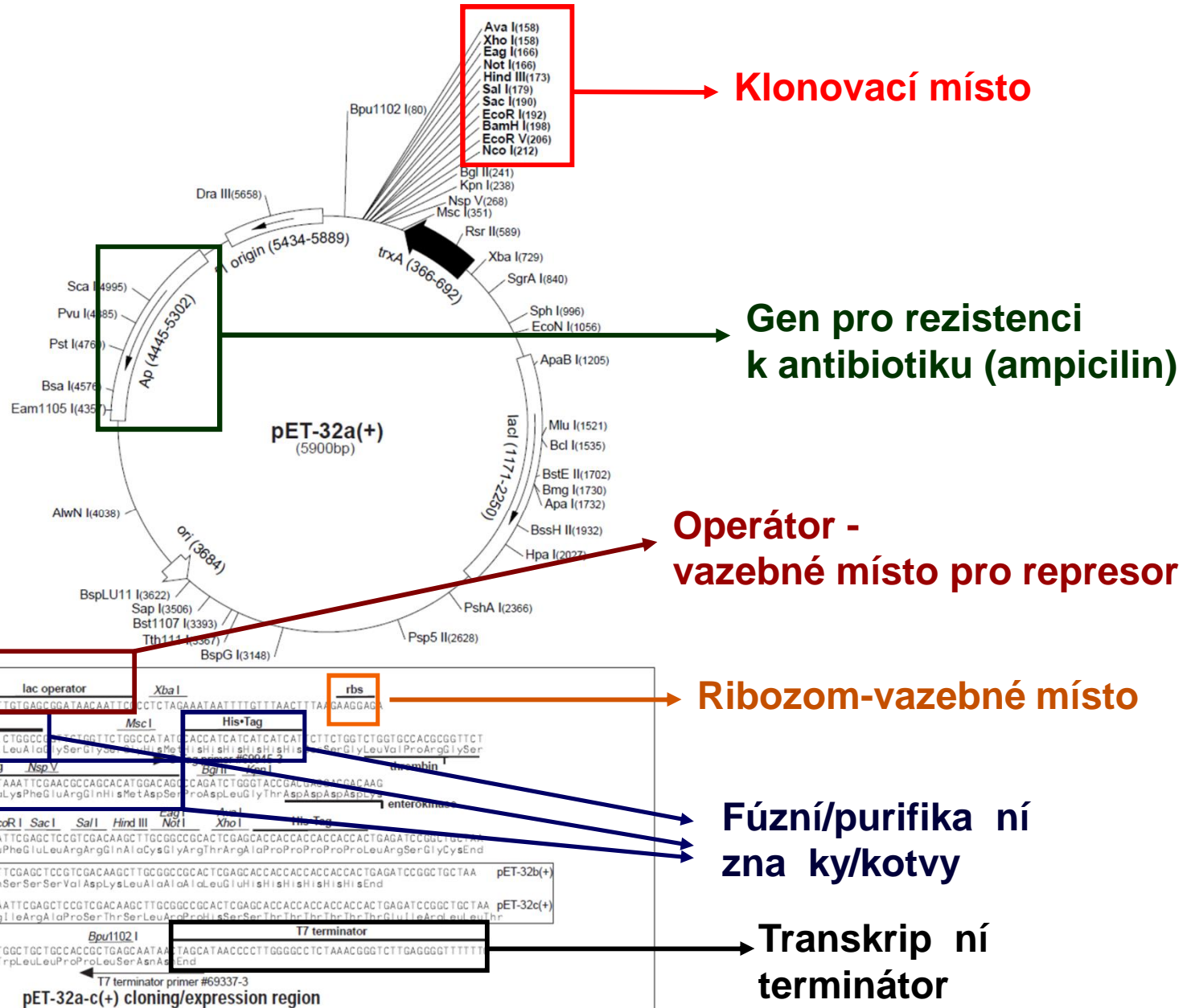


# Struktura vektoru pro ú innou expresi rekombinantních protein v *E.coli*

## pET-32a(+) sequence landmarks

T7 promoter	764-780
T7 transcription start	763
Trx*Tag coding sequence	366-692
His*Tag coding sequence	327-344
S*Tag coding sequence	249-293
Multiple cloning sites ( <i>Nco</i> I - <i>Xho</i> I)	158-217
His*Tag coding sequence	140-157
T7 terminator	26-72
<i>lacI</i> coding sequence	1171-2250
pBR322 origin	3684
<i>bla</i> coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209.

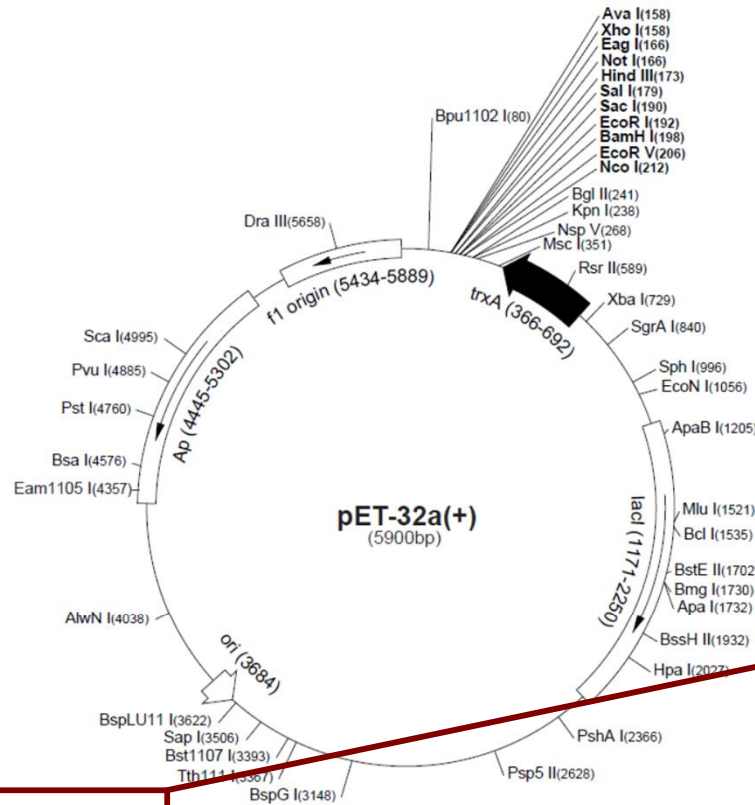


# Struktúra vektoru pro ú innou expresi rekombinantních proteinů v *E.coli*

## pET-32a(+) sequence landmarks

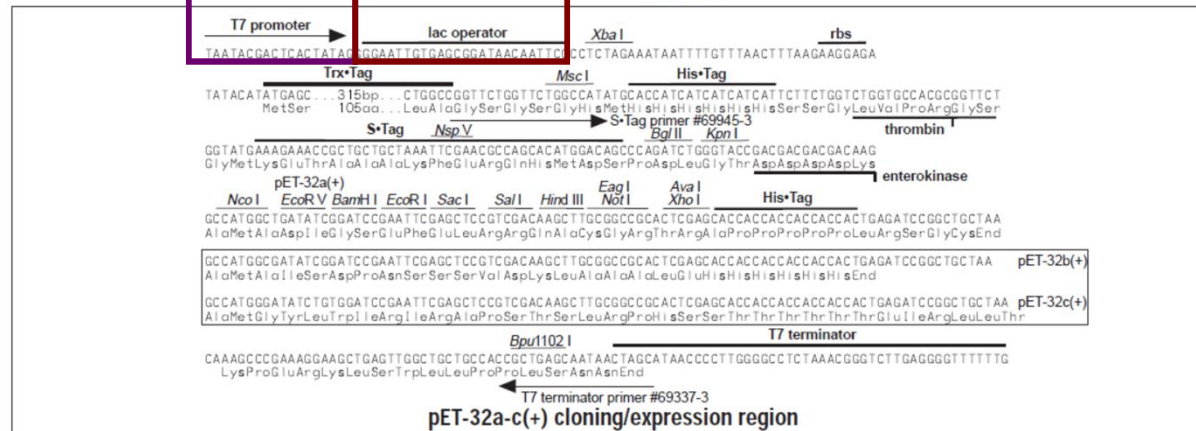
T7 promoter	764-780
T7 transcription start	763
Trx*Tag coding sequence	366-692
His*Tag coding sequence	327-344
S*Tag coding sequence	249-293
Multiple cloning sites ( <i>Nco</i> I - <i>Xho</i> I)	158-217
His*Tag coding sequence	140-157
T7 terminator	26-72
<i>lacI</i> coding sequence	1171-2250
pBR322 origin	3684
<i>bla</i> coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209.



promotor

Operátor -  
vazebné místo pro represor



## **Vlastnosti promotoru:**

**ÉSilný promotor** (ptac, ptrp,  $\lambda$ pL, pT<sub>7</sub>)

- Protein zájmu by m l tvo it 10-30% a více z celkového bakteriálního proteinu.

**ÉP enositelný do r zných kmen *E.coli***

**ÉJednodu-e a levn inducibilní**

- Teplotní indukce ( $\lambda$ pL)

- Chemická indukce (ptac, ptrp, pT7): IPTG (isopropyl- $\beta$ -D-thiogalaktopyranozid)

**ÉVykazuje minimální hladinu bazální exprese**

- Pokud jsou proteiny netoxické dosahuje se vysokých výt flk protein r stem bun k do vysokých hustot s následnou indukcí aktivity promotoru.

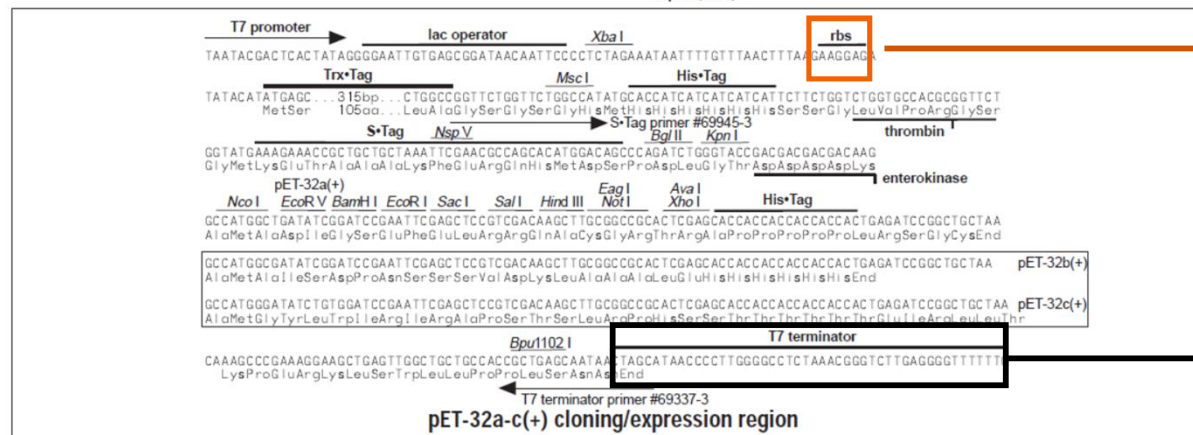
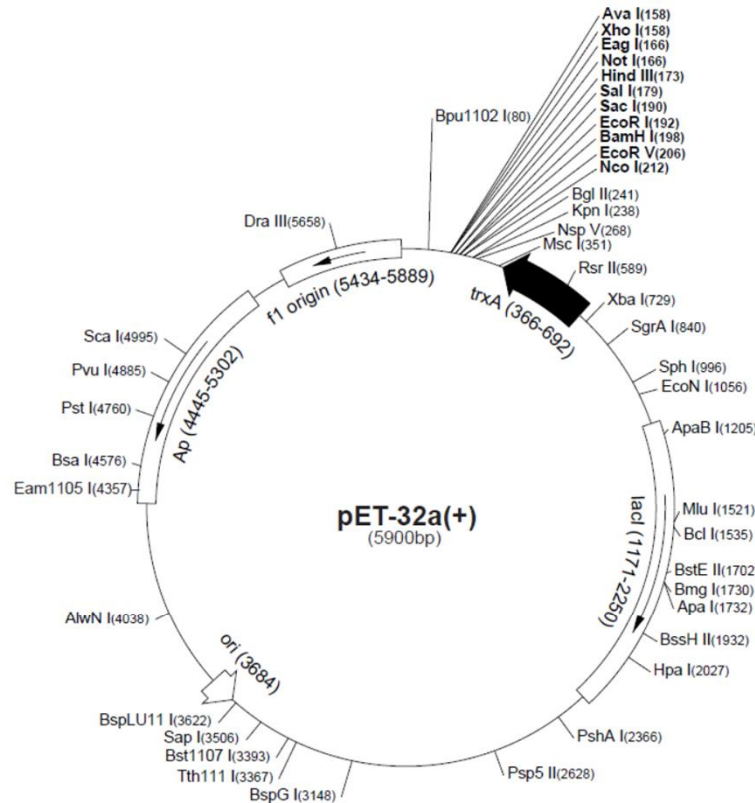
- U toxických protein pro je nutná minimalizace bazální transkripce p ed p ídavkem induk ního inidla pomocí vhodného represoru.

# Struktura vektoru pro ú innou expresi rekombinantních proteinů v *E.coli*

## pET-32a(+) sequence landmarks

T7 promoter	764-780
T7 transcription start	763
Trx*Tag coding sequence	366-692
His*Tag coding sequence	327-344
S*Tag coding sequence	249-293
Multiple cloning sites ( <i>Nco</i> I - <i>Xho</i> I)	158-217
His*Tag coding sequence	140-157
T7 terminator	26-72
<i>lac</i> I coding sequence	1171-2250
pBR322 origin	3684
<i>bla</i> coding sequence	4445-5302
f1 origin	5434-5889

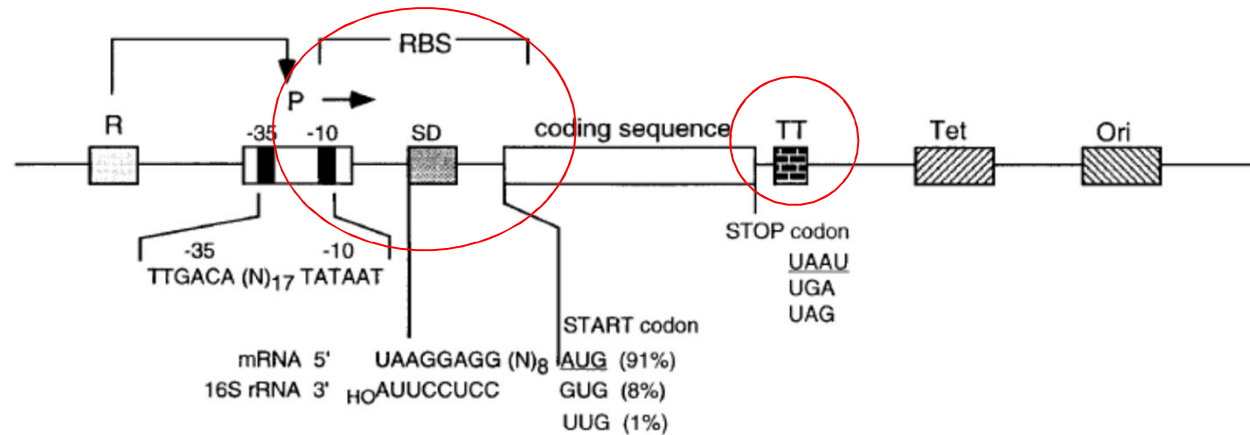
The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*HI at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*HI at 198 except for *Eco*R V, which cuts at 209.



Ribozom-vazebné místo

Transkripční terminátor

# Struktura vektoru pro účinnou expresi rekombinantních proteinů v *E.coli*



**Ribozom-vazebné místo:** zahrnuje Shine-Dalgarnova (SD) sekvenci a translační iniciační kodon

Vzdálenost mezi SD sekvencí a iniciačním kodonem AUG je 4-13 nukleotidů, tato vzdálenost ovlivňuje účinnost iniciace translace (**optimální vzdálenost je 4-8 nukleotidů**), oblast bohatá na AT páry.

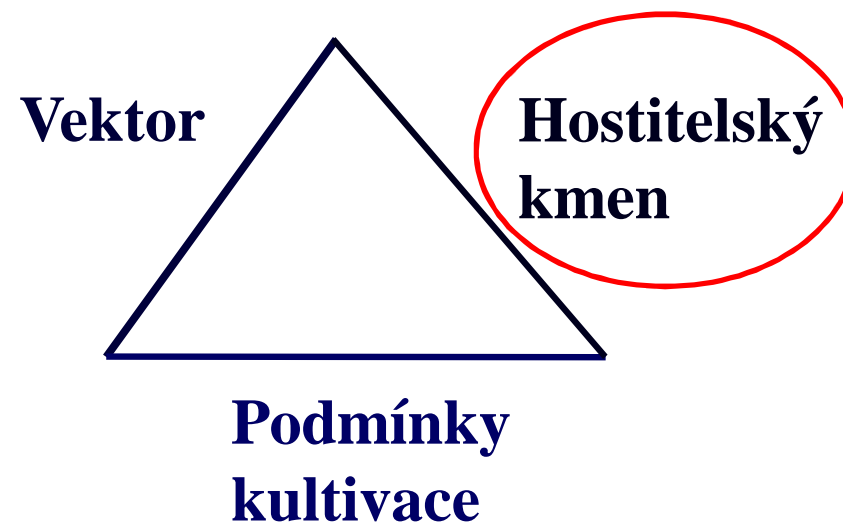
**Transkripční terminátor**

T<sub>7</sub> term, rrnT1, T2

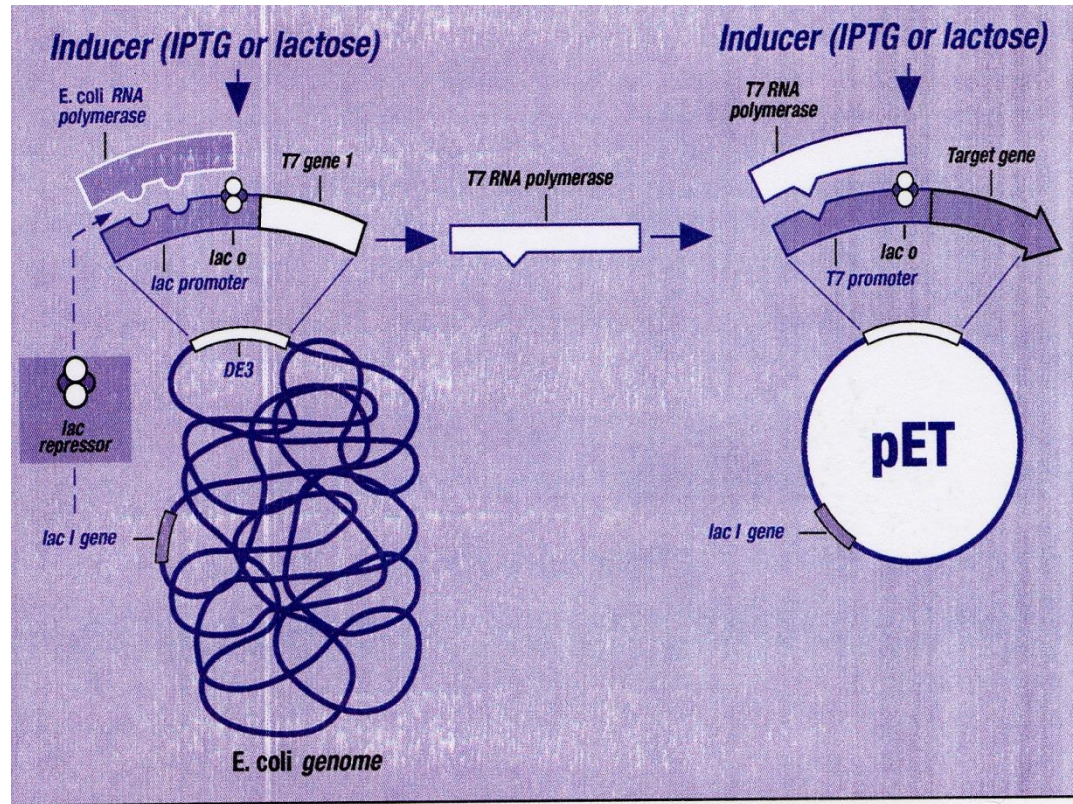
(zabraňuje okluzi promotoru, zvyšuje stabilitu mRNA)



# Výběr hostitelského kmene *E.coli*



# Expresa rekombinantního proteinu v hostitelském kmeni *E. coli* BL21(DE3)



## **Toxicita rekombinantního proteinu pro hostitelský kmen**

Nejen omezena na pouhý fakt, že je protein je pro buňku cizí, ale může být způsobena i tím, že je nadprodukován určitý nativní gen.

### **Pro hostitele jsou smrtelné:**

Rekombinantní proteiny s hydrofóbními oblastmi mají toxický účinek - asociují s membránou nebo se inkorporují do membránového systému buňky (porušení membránového potenciálu).

Proteiny, které inaktivují ribozomy.

## Výběr hostitelského kmene *E.coli* s ohledem na toxicitu proteinu pro hostitele



ÉNutná p ísná regulace expresního systému

Komer n dostupné bakteriální kmeny s r znými úrovn ěmi regulace exprese, zaji– ujícími minimalizaci bazální exprese.

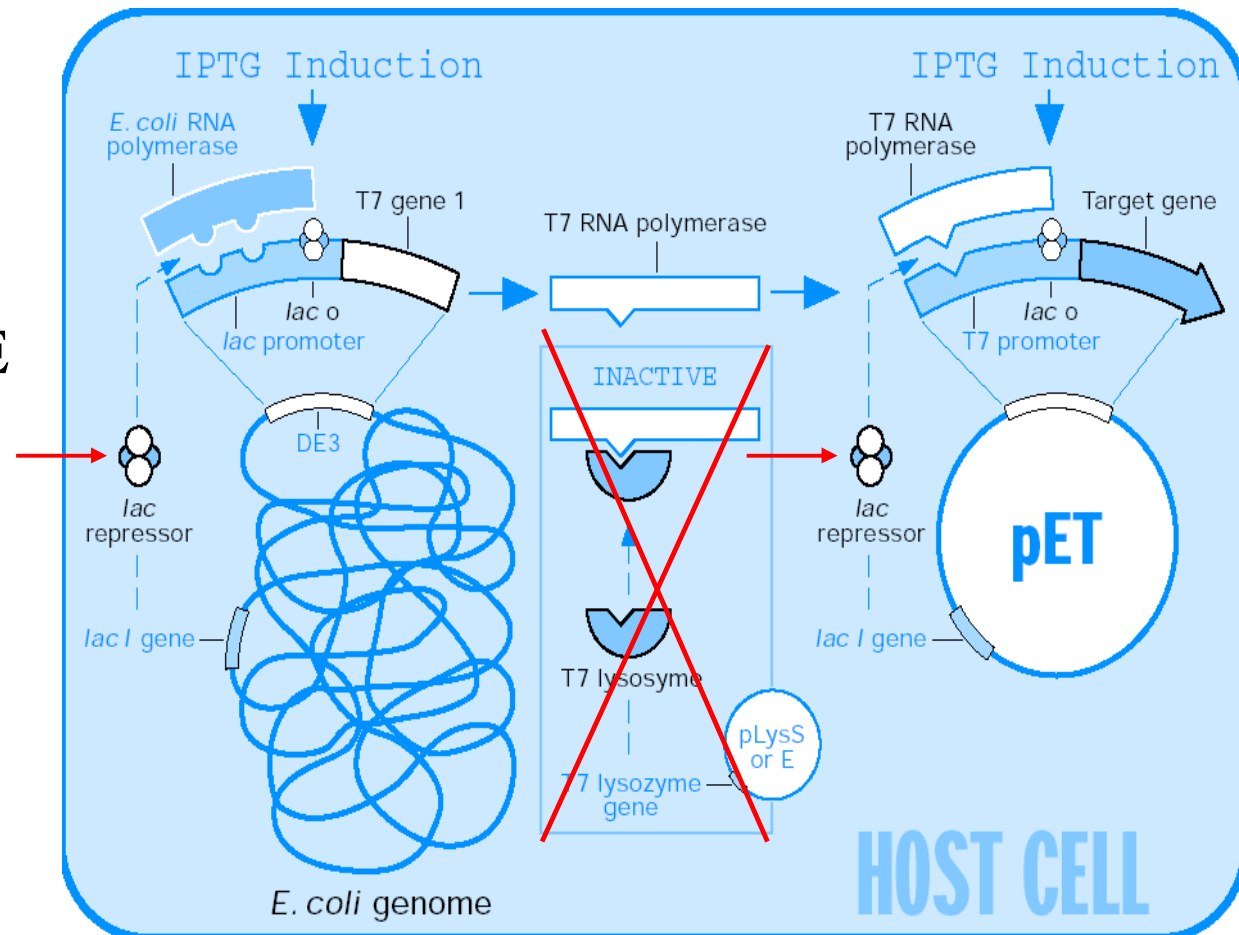
<b>BL21(DE3)</b>	firma Novagen
<b>BL21(DE3)pLysS</b>	firma Novagen

# Různé úrovně minimalizace bazální exprese

✓ **BL21(DE3)**

BL21(DE3)pLysS/E

BL21



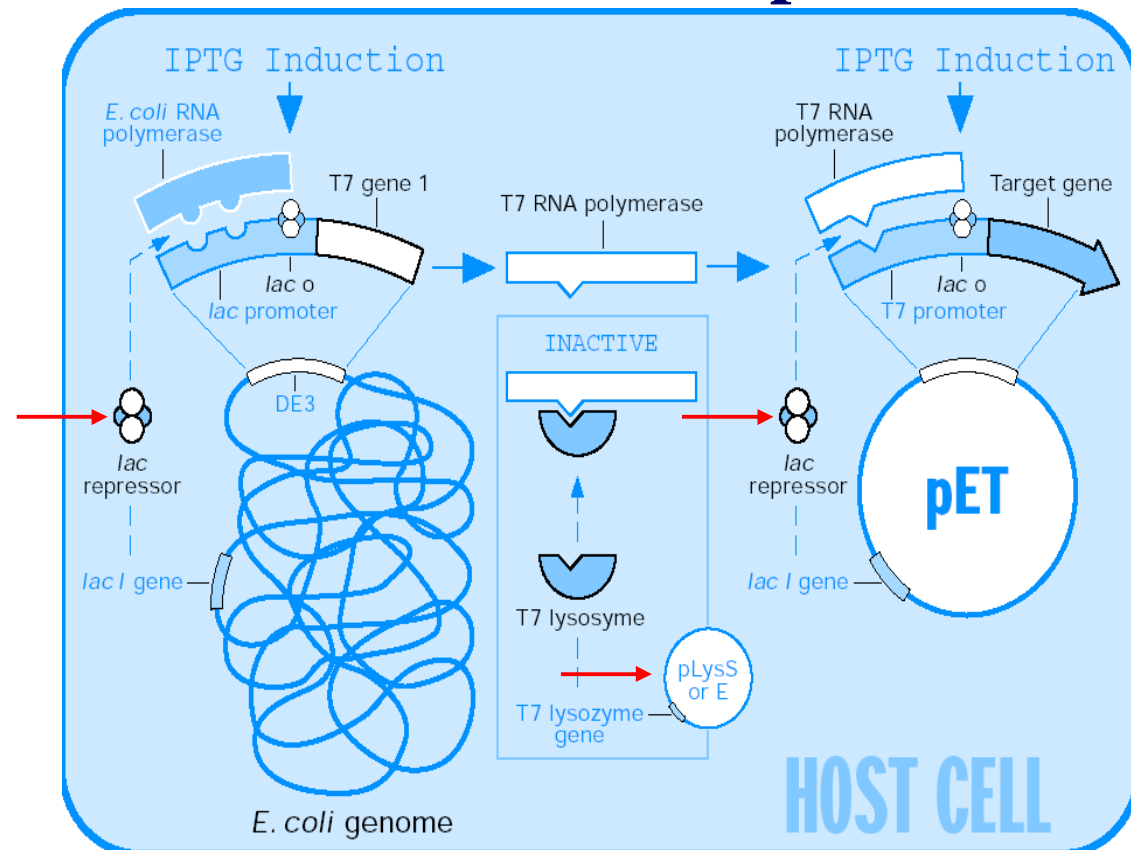
Cca 10 % hladina bazální exprese (před indukcí exprese) klonovaného genu.

## Různé úrovně minimalizace bazální exprese

BL21(DE3)

✓ **BL21(DE3)pLysS/E**

BL21



Expresní kmen obsahující plazmidy pLysS nebo pLysE umožní účinnou regulaci expresního systému využívající T7 promotor. Tyto plazmidy kódují lysozym, který se váže na T7 RNA polymerasu a inaktivuje ji a minimalizuje se tak bazální exprese.

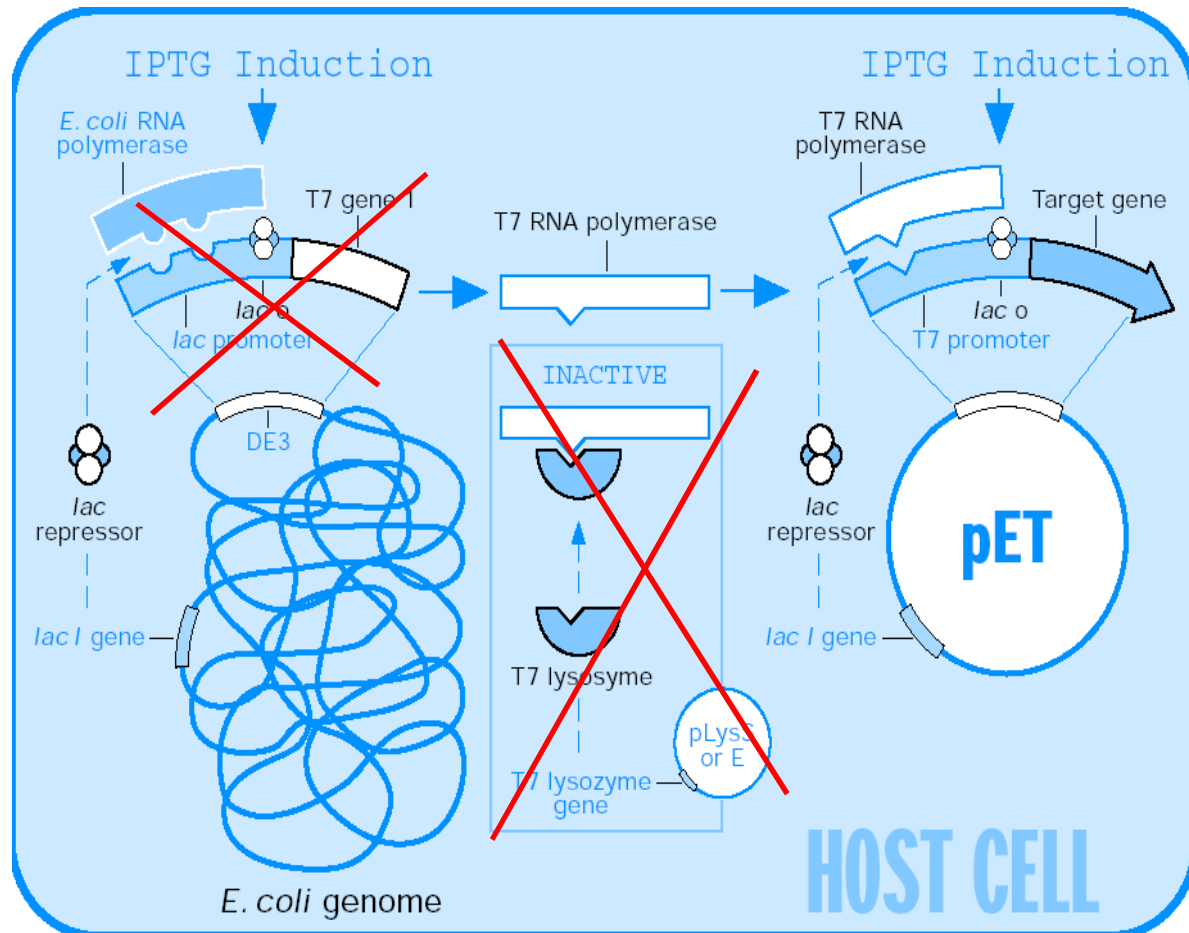
Cca 1-3% hladina bazální (před indukci exprese) exprese klonovaného genu.

# R zné úrovn minimalizace bazální exprese

BL21(DE3)

BL21(DE3)pLysS/E

✓ **BL21**



ÉIndukce exprese infekcí bakteriofágem CEG (gen pro T7 RNA polymerasu)

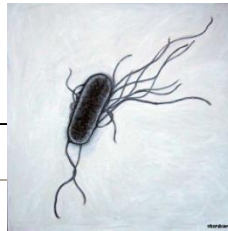
Nejvyší úroveň exprese!!

# Využívání kodon *E.coli* (codon usage)

ÉGeny u prokaryot a eukaryot se vyzna ují nenáhodným využíváním synonymních kodon .

ÉKodony z ídka využívané u *E. Coli* se mohou hojn vyskytovat u heterologních gen pocházejících z eukaryot, archebakterií aj.

ÉFrekvence využití synonymních kodon obvykle odráží zastoupení jejich tRNA v cytoplazm .



<i>Escherichia coli</i> K12 [gbbet]: 14 CDS's (5122 codons)								<i>Arabidopsis thaliana</i> [gbpln]: 80395 CDS's (31098475 codons)							
fields: [triplet] [frequency: per thousand] ([number])								fields: [triplet] [frequency: per thousand] ([number])							
UUU 19.7 ( 101)	UCU 5.7 ( 29)	UAU 16.8 ( 86)	UGU 5.9 ( 30)	UUC 15.0 ( 77)	UCC 5.5 ( 28)	UAC 14.6 ( 75)	UGC 8.0 ( 41)	UUU 21.8 (678320)	UCU 25.2 (782818)	UAU 14.6 (455089)	UGU 10.5 (327640)	UUC 20.7 (642407)	UCC 11.2 (348173)	UAC 13.7 (427132)	UGC 7.2 (222769)
UUA 15.2 ( 78)	UCA 7.8 ( 40)	UAA 1.8 ( 9)	UGA 1.0 ( 5)	UUA 12.7 (394867)	UCA 18.3 (568570)	UAA 0.9 ( 29405)	UGA 1.2 ( 36260)	UUA 12.7 (394867)	UCA 18.3 (568570)	UAA 0.9 ( 29405)	UGA 1.2 ( 36260)	UUA 12.7 (394867)	UCA 18.3 (568570)	UAA 0.9 ( 29405)	UGA 1.2 ( 36260)
UUG 11.9 ( 61)	UCG 8.0 ( 41)	UAG 0.0 ( 0)	UGG 10.7 ( 55)	UUG 11.9 ( 61)	UCG 8.0 ( 41)	UAG 0.0 ( 0)	UGG 10.7 ( 55)	UUG 20.9 (649150)	UCG 9.3 (290158)	UAG 0.5 ( 16417)	UGG 12.5 (388049)	UUG 20.9 (649150)	UCG 9.3 (290158)	UAG 0.5 ( 16417)	UGG 12.5 (388049)
CUU 11.9 ( 61)	CCU 8.4 ( 43)	CAU 15.8 ( 81)	CGU 21.1 ( 108)	CUC 10.5 ( 54)	CCC 6.4 ( 33)	CAC 13.1 ( 67)	CGC 26.0 ( 133)	CUU 24.1 (750114)	CCU 18.7 (580962)	CAU 13.8 (428694)	CGU 9.0 (280392)	CUC 16.1 (500524)	CCC 5.3 (165252)	CAC 8.7 (271155)	CGC 3.8 (117543)
CUA 5.3 ( 27)	CCA 6.6 ( 34)	CAA 12.1 ( 62)	CGA 4.3 ( 22)	CUA 5.3 ( 27)	CCA 6.6 ( 34)	CAA 12.1 ( 62)	CGA 4.3 ( 22)	CUA 9.9 (307000)	CCA 16.1 (502101)	CAA 19.4 (604800)	CGA 6.3 (195736)	CUA 9.9 (307000)	CCA 16.1 (502101)	CAA 19.4 (604800)	CGA 6.3 (195736)
CUG 46.9 ( 240)	CCG 26.7 ( 137)	CAG 27.7 ( 142)	CGG 4.1 ( 21)	CUG 46.9 ( 240)	CCG 26.7 ( 137)	CAG 27.7 ( 142)	CGG 4.1 ( 21)	CUG 9.8 (305822)	CCG 8.6 (268115)	CAG 15.2 (473809)	CGG 4.9 (151572)	CUG 9.8 (305822)	CCG 8.6 (268115)	CAG 15.2 (473809)	CGG 4.9 (151572)
AUU 30.5 ( 156)	ACU 8.0 ( 41)	AAU 21.9 ( 112)	AGU 7.2 ( 37)	AUC 18.2 ( 93)	ACC 22.8 ( 117)	AAC 24.4 ( 125)	AGC 16.6 ( 85)	AUU 21.5 (668227)	ACU 17.5 (544807)	AAU 22.3 (693344)	AGU 14.0 (435738)	AUC 18.5 (576287)	ACC 10.3 (321640)	AAC 20.9 (650826)	AGC 11.3 (352568)
AUA 3.7 ( 19)	ACA 6.4 ( 33)	AAA 33.2 ( 170)	AGA 1.4 ( 7)	AUA 3.7 ( 19)	ACA 6.4 ( 33)	AAA 33.2 ( 170)	AGA 1.4 ( 7)	AUA 12.6 (391867)	ACA 15.7 (487161)	AAA 30.8 (957374)	AGA 19.0 (589788)	AUA 12.6 (391867)	ACA 15.7 (487161)	AAA 30.8 (957374)	AGA 19.0 (589788)
AUG 24.8 ( 127)	ACG 11.5 ( 59)	AAG 12.1 ( 62)	AGG 1.6 ( 8)	AUG 24.8 ( 127)	ACG 11.5 ( 59)	AAG 12.1 ( 62)	AGG 1.6 ( 8)	AUG 24.5 (762852)	ACG 7.7 (240652)	AAG 32.7 (1016176)	AGG 11.0 (340922)	AUG 24.5 (762852)	ACG 7.7 (240652)	AAG 32.7 (1016176)	AGG 11.0 (340922)
GUU 16.8 ( 86)	GCU 10.7 ( 55)	GAU 37.9 ( 194)	GGU 21.3 ( 109)	GUC 11.7 ( 60)	GCC 31.6 ( 162)	GAC 20.5 ( 105)	GGC 33.4 ( 171)	GUU 27.2 (847061)	GCU 28.3 (880808)	GAU 36.6 (1139637)	GGU 22.2 (689891)	GUC 12.8 (397008)	GCC 10.3 (321500)	GAC 17.2 (535668)	GGC 9.2 (284681)
GUA 11.5 ( 59)	GCA 21.1 ( 108)	GAA 43.7 ( 224)	GGA 9.2 ( 47)	GUA 11.5 ( 59)	GCA 21.1 ( 108)	GAA 43.7 ( 224)	GGA 9.2 ( 47)	GUA 9.9 (308605)	GCA 17.5 (543180)	GAA 34.3 (1068012)	GGA 24.2 (751489)	GUA 9.9 (308605)	GCA 17.5 (543180)	GAA 34.3 (1068012)	GGA 24.2 (751489)
GUG 26.4 ( 135)	GCG 38.5 ( 197)	GAG 18.4 ( 94)	GGG 8.6 ( 44)	GUG 26.4 ( 135)	GCG 38.5 ( 197)	GAG 18.4 ( 94)	GGG 8.6 ( 44)	GUG 17.4 (539873)	GCG 9.0 (280804)	GAG 32.2 (1002594)	GGG 10.2 (316620)	GUG 17.4 (539873)	GCG 9.0 (280804)	GAG 32.2 (1002594)	GGG 10.2 (316620)
Coding GC 52.35% 1st letter GC 60.82% 2nd letter GC 40.61% 3rd letter GC 55.62%								Coding GC 44.59% 1st letter GC 50.84% 2nd letter GC 40.54% 3rd letter GC 42.38%							

<http://www.kazusa.or.jp/codon/>



## Málo preferované kodony u *E.coli*

Codon(s)	Amino acid
AGA, AGG, CGA, CGG.....	Arg
UGU, UGC.....	Cys
GGA, GGG.....	Gly
AUA.....	Ile
CUA, CUC.....	Leu
CCC, CCU, CCA.....	Pro
UCA, AGU, UCG, UCC .....	Ser
ACA .....	Thr

Makrides, 1996

Expres heterologní gen obsahující málo preferované kodony vede k translačním chybám !

ÉP ed asné ukonění translace (zkrácený produkt)

ÉPosunutí tečího rámce (posun aflu 2 AK v místě kodonu AGA)

ÉZámna aminokyseliny - místo arginin (kodon AGA) za lyzin

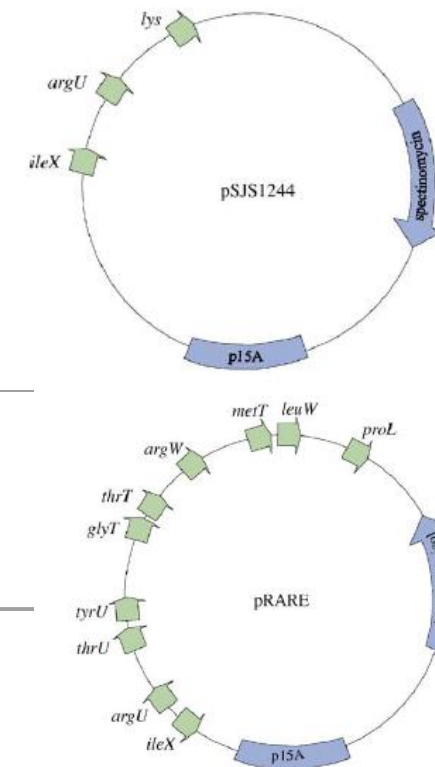
## Výběr hostitelského kmene *E.coli*



Pro produkci genů obsahující velké množství kodonů, které jsou málo využívané *E.coli*, je možné použít komerčně dodávané kmeny, které produkují tRNA málo užívaných kodonů.

Plasmidy komplementující tRNA.

<p>BL21 (DE3) CodonPlus-RIL</p>	<p>AGG/AGA (arginine, <b>R</b>), AUA (isoleucine, <b>I</b>) and CUA (leucine, <b>L</b>)</p>	<p>firma Stratagene</p>
<p>BL21 (DE3) CodonPlus-RP</p>	<p>AGG/AGA (arginine, <b>R</b>) and CCC (proline, <b>P</b>)</p>	<p>firma Stratagene</p>
<p>Rosetta or Rosetta (DE3)</p>	<p>AGG/AGA (arginine, <b>R</b>), CGG (arginine, <b>R</b>), AUA (isoleucine, <b>I</b>) CUA (leucine, <b>L</b>) CCC (proline), and GGA (glycine, <b>G</b>)</p>	<p>firma Novagen</p>



NEBO: Místní řízená mutagenese - zaměřena na málo využívaného kodonu.

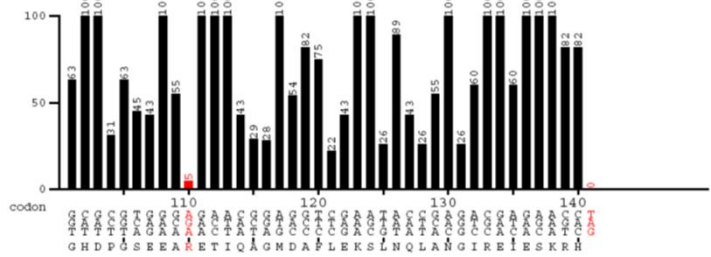
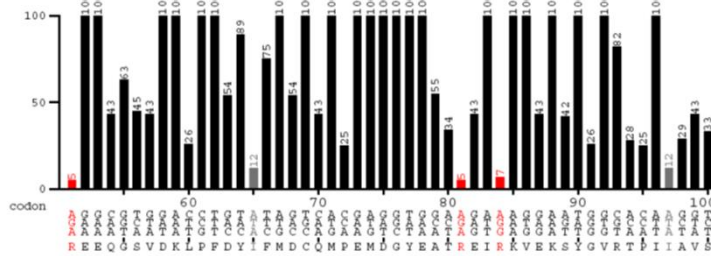
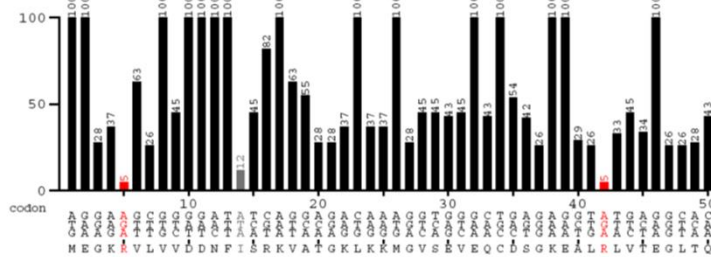
# Optimalizace kodon ó syntéza gen

Optimální složení kodon pro produkci v jednom i více organismech s ohledem na vznik sekundárních struktur a stabilitu mRNA.



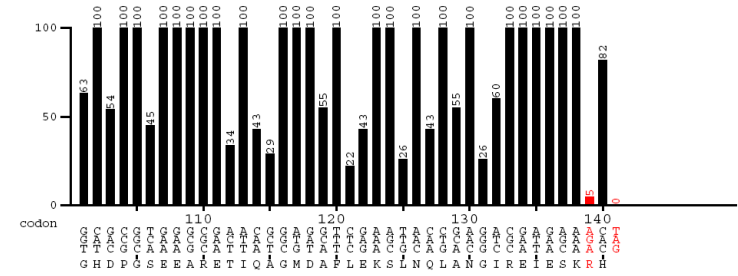
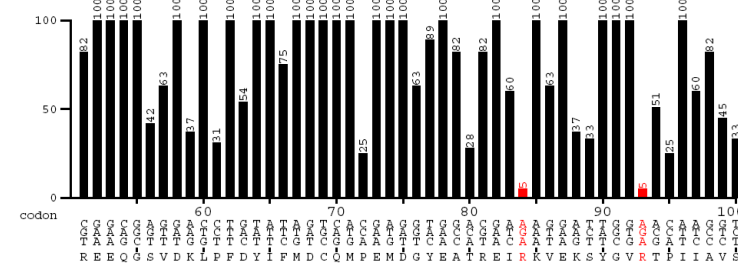
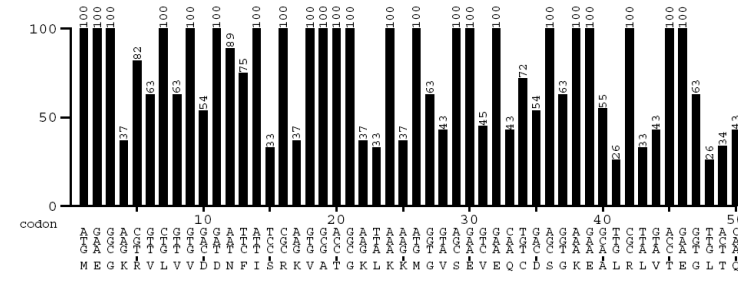
www.gcu.a.de  
created: 18.08.2011

CKIrd  
sequence derived from Arabidopsis\_thaliana  
Codontable:  
Escherichia\_coli\_K12  
Ordinate (y-axis): relative adaptiveness <20% <10%



www.gcu.a.de  
created: 24.08.2011

CKIrd optim codon us  
sequence derived from Arabidopsis\_thaliana  
Codontable:  
Escherichia\_coli\_K12  
Ordinate (y-axis): relative adaptiveness <20% <10%



# Degradace proteinu

## Bakteriální proteolytický systém:

- *E. coli* obsahuje velké množství proteas, nejvíce v cytoplazm

- Selektivn odstra uje šabnormálníõ proteiny:

ÉNekompletní polypeptidy

ÉProteiny se zam n nými AK

ÉNadm rn syntetizované podjednotky multimerních protein

ÉProteiny po-kozené oxidací nebo volnými radikály

ÉCizí, rekombinantní proteiny (problémem jsou proteiny < 10 kDa)

# Aminokyseliny redukující stabilitu heterologního proteinu

## 1. N-koncové pravidlo

É Stabilita proteinu je ovlivněna aminokyselinou, která následuje první aminokyselinou polypeptidového řetězce (methionin).

É Aminokyseliny na této pozici redukují stabilitu proteinu:

**Arg, Lys, Leu, Phe, Tyr a Trp**

É Poločas rozpadu proteinu pouze 2 min

## 2. Lysin ve vnitřní sekvenci poblíž N-konce proteinu

É Degradace ubiquitin-dependentními proteasami

## 3. PEST hypotéza

É Oblasti bohaté na Pro (P), Glu (E), Ser (S), Thr (T)

É Po fosforylaci PEST degradace proteinu Ca-dependentními proteasami

## Výběr hostitelského kmene *E.coli*



### Kmeny deficientní na proteasy

ÉMutace, eliminující produkci proteas a tím proteolytickou degradaci rekombinantních proteinů .

Expresní kmeny **BL21** jsou deficientní na:

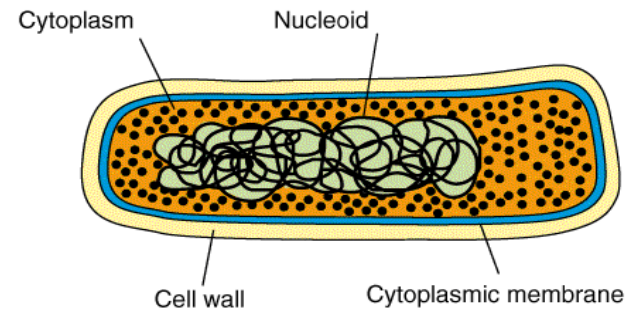
cytoplazmatickou proteasu *lon*  
periplazmatickou proteasu *ompT*

# Cílená exprese proteinu

**Cytoplazmatická exprese**

**Periplazmatická exprese**

**Extracelulární exprese  
(do kultiva ního média)**



# Cytoplazmatická exprese

ÉPreferovaný zp sob

## Výhody

ÉVysoký výt flek proteinu

ÉJednoduší plazmidové konstrukty

ÉInkluzní t líska

## Nevýhody

ÉInkluzní t líska

ÉReduk ní prost edí

ÉProteolýza

ÉVíce komplexní purifikace



# Inkluzní tělíska

Nerozpustné shluky (cca  $2\mu\text{m}^3$ ) složené z nativních proteinů s nízkou rozpustností, z nesložených nebo z částečně poskládaných proteinů

## Co způsobuje jejich tvorbu?

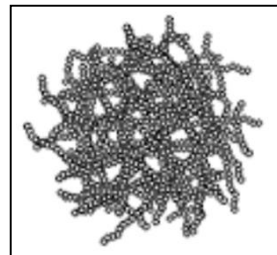
1. Prostředí v *E. coli* se liší od přirozeného prostředí ve smyslu redoxního potenciálu (redukční prostředí v cytoplazmě *E. coli*), pH, osmolarita, absence chaperonů, kofaktorů, absence post-translačních modifikací

2. Vysoká hladina exprimovaných proteinů

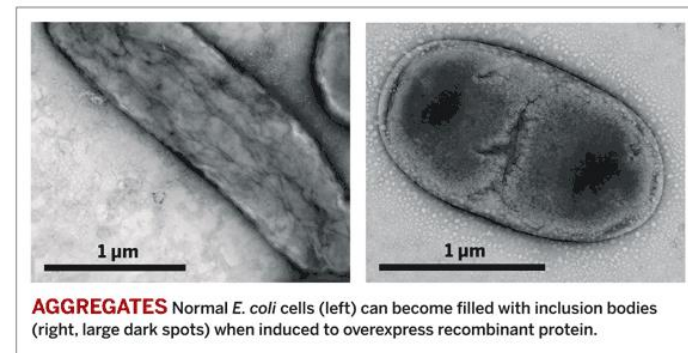
- exponované hydrofobní domény nesbalených polypeptidových řetězců navzájem (intramolekulárně) asociují



Rozpustný protein



Inkluzní tělíska z nerozpustného proteinu



# Inkluzní t líska

## Výhody

ÉSnadná izolace ve vysoké čistotě a koncentraci

ÉOchrana před proteasami

ÉPro produkci proteinů, jejichž aktivita je pro buňku letální

## Nevýhody

ÉProteinová nerozpustnost

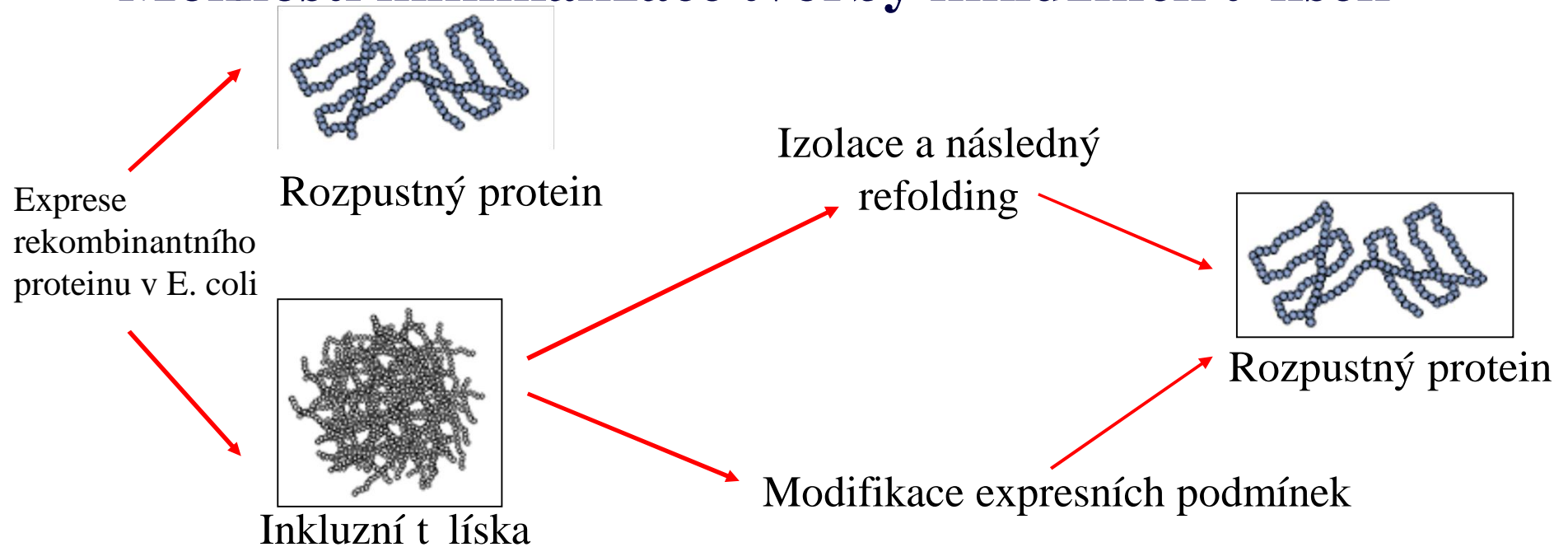
ÉRefolding pro opětovné získání biologické aktivity

ÉRefolding nemusí vést k zaktivování proteinu

ÉRedukce výtěžku proteinu

ÉZvyšují se náklady

# Možnosti minimalizace tvorby inkluzních tělíček



Například:

• Snížení teploty kultivace bakteriální kultury

• Koprodukce chaperon

• Použití fúzního partnera zlepšujícího solubilizaci (thioredoxin)

• Selektce hostitele - například bakteriální kmene deficientní na thioredoxin reduktasu

# Možnosti minimalizace tvorby inkluzních tělíček

## Výběr hostitelského kmene *E.coli*



Pokud protein obsahuje jeden nebo více disulfidických vazeb, správné poskládání proteinu je stimulováno v hostiteli s více oxidujícím prostředím cytoplazmy.

AD494	Mutace v genu pro thioredoxinreduktasu ( <b>trxB</b> )	Novagen
Origami	Dvojitá mutace v genu pro thioredoxinreduktasu ( <b>trxB</b> ) and glutathionreduktasu ( <b>gor</b> )	Novagen

# Periplazmatická exprese

É Periplazma obsahuje jen 4% všech buněčných proteinů (cca 100 proteinů)

É Transmembránový transport je zprostředkován signálním peptidem na N-konci proteinu

É Prokaryotické signální peptidy úspěšně použité v *E.coli* (ompA, ompT z *E.coli*, protein A z *S. Aureus*, endoglukanasa z *B.subtilis*)

## Výhody

É Jednodušší purifikace

É Není zde tak rozsáhlá proteolýza

É Zlepšení tvorby disulfidických můstků /foldingu

## Nevýhody

É Signální peptid nezajistí vždy transport do periplazmy

É Mohou se také tvořit inkluzní tělíska

# Extracelulární exprese

É Sekrece proteinů do kultivačního média

É Chybí účinná cesta pro transport skrz vnější membránu (*E. coli* sekretuje velmi málo proteinů).

É Některé proteiny sekretované do periplazmy pasivně difundují do média.

É Zatím spíše neúspěšná manipulace s různými transportními cestami, které by usnadnily sekreci cizího proteinu.

## Výhody

É Minimální kontaminace ostatními proteiny (jednodušší purifikace)

É Nejmenší hladina proteolýzy

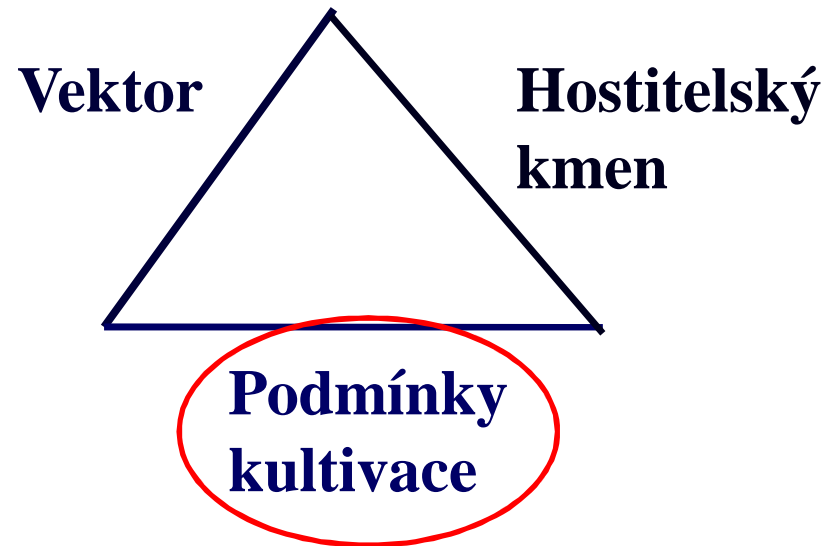
É Zlepšení foldingu

## Nevýhody

É Často nízká sekrece

É Hodně zedarovaný protein

# Modifikace r stových podmínek



# Podmínky kultivace

Možnosti zvýšení produkce rozpustného proteinu:

Experimentálně se testuje:

É Vysoká hustota buněčné kultury

É Složení média (pH, přísady specifických substrátů, kofaktorů, složení živin-bohaté i minimální média)

É Optimalizace teploty růstu bakterií a teploty na indukci exprese

É Koncentrace indukčního inidla na indukci, délka indukce exprese



## Podmínky kultivace - složení média

Vliv různých podmínek exprese na aktivitu mutantní formy kukuřičné  $\beta$ -glukosidasy F461L.

ÉpH média

ÉP ítomnost substrátu  
(celobiosa)

<i>podmínky</i>	<i>Specifická aktivita (nkat/mg)</i>
kontrola	1,9
LB médium pH 6	2,0
LB médium pH 7	4,2
LB médium pH 8	2,8
1% celobiosa (na indukci exprese)	2,7

Specifická aktivita byla po expresi za uvedených podmínek měřena v nativních lyzátech použitím substrátu PNPG.

# Expresse AHP protein v *E. coli* o optimalizace teploty r stu bakterií a teploty na indukci exprese

## Podmínky:

1. Médium: LB médium, bakteriální kmen BL21(DE3)pLysS

*R st* ( $OD_{600} \sim 0.5-0.6$ )

Indukce 0,4 mM IPTG/3hodiny

22°C

22°C (3)

1

2

3

37°C

22°C (2)

37°C

28°C (1)

## Test rozpustnosti:

2. Příprava proteinových lyzátů za silných denaturačních podmínek (celková produkce proteinu = rozpustná + nerozpustná forma) a za nativních podmínek (rozpustná forma proteinu).

3. SDS PAGE denuračních a nativních proteinových lyzátů s následnou analýzou western blottingem.

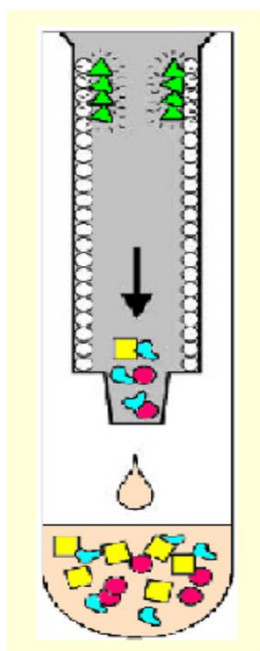


4. Detekce proteinu pomocí série protilátek a kvantifikace signálů pomocí programu pro analýzu 1-D gelů (př. Quantity One- BioRad, Quanti Scan-přístupný na internetu).

## Exprese AHP protein v *E. coli* -optimalizace teploty r stu bakterií a indukce exprese

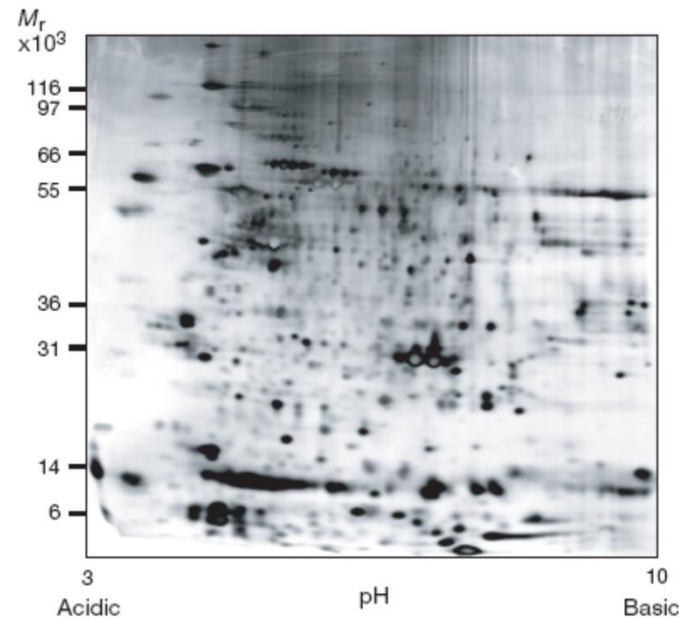
Procenta produkce AHP protein v rozpustné form						
<i>t</i> (°C) <i>R st</i> /indukce	AHP1	AHP2	AHP3	AHP4	AHP5	AHP6
37°C/28°C	8 %	85 %	100%	0	76 %	0
37°C/22°C	82 %	73 %	100%	0	81 %	51 %
22°C/22°C	71 %	78 %	100%	30 %	81 %	73 %

## 2. část: Purifikace rekombinantních protein



Purifikace protein fúzovaných s GFP  
pomocí hydrofóbní matrice

# Purifikace proteinu z komplexní směsi makromolekul přítomných v biologickém vzorku (buněčný nebo tkáňový extrakt)



Analýza buněčného extraktu 2D elektroforézou

ÉN kolik tisíc proteinů s různými vlastnostmi (~5000-8000) a v různých množstvích (aktin ~ 10%, unikátní transkripční faktor < 0,001% z celkového proteinu)

ÉDNA, RNA, polysacharidy, lipidy

*Neflza nemeí í í í í í í í*

***1. Pro ???***

**Pro jaký účel ?**

***2. Jak???***

**Jak protein detekovat?**

***3. Co???***

**Jaké vlastnosti má protein ?**

# 1. Pro ???

# Pro jaký účel ?

<b>Aplikace</b>	<b>Množství</b>	<b>čistota</b>	<b>Poznámka</b>
<b>Identifikace</b>	<b>0,002-0,2 µg</b>	<b>vysoká (&gt;95%)</b>	ÉEdmanovo odbourávání (5-10 pmol), přístupy hmotnostní spektroskopie (0,2-1 pmol)
<b>Produkce protilátek</b>	<b>µg-mg</b>	<b>střední-vysoká</b>	ÉPro imunizaci: ~ 0,1 µg proteinu É čím větší čistota tím větší a rychlejší –ance pro získání vysoce specifické imunitní odpovědi.
<b>Enzymologie</b>	<b>1-5 mg</b>	<b>vysoká &gt; 95 %</b>	ÉMnožství proteinu závisí na citlivosti analýzy. É čistota závisí na specifitě analýzy a ovlivňuje výsledky analýzy kontaminacemi.
<b>Biofyzikální studie</b>	<b>mg-g</b>	<b>vysoká (&gt;95%)</b>	ÉCD spektroskopie, rezonance povrchových plasmonů, fluorimetrie, analytická ultracentrifugace, UV spektroskopie
<b>3D struktura (krystalizace, NMR)</b>	<b>10-20 mg</b>	<b>vysoká (&gt;95%)</b>	ÉHledání krystalizačních podmínek ~ 1-2 mg proteinu, získání krystalu o velikosti dostatečné pro rentgenovou difrakci 5-10 mg proteinu ÉPro první 1-D NMR spektra se vyžaduje ~ 0,5 µmol proteinu, protein (velikost 5-20kDa) značený <sup>15</sup> N / <sup>13</sup> C je nutný pro vyřešení struktury.
<b>Farmaceutické účely</b>	<b>mg-kg</b>	<b>vysoká (99,9%)</b>	ÉPro klinické účely nesmí proteiny obsahovat pyrogeny a bakteriální toxiny a musí být velmi stabilní kvůli dlouhodobému skladování.

## 2. *Jak???*

# Jak budeme protein analyzovat?

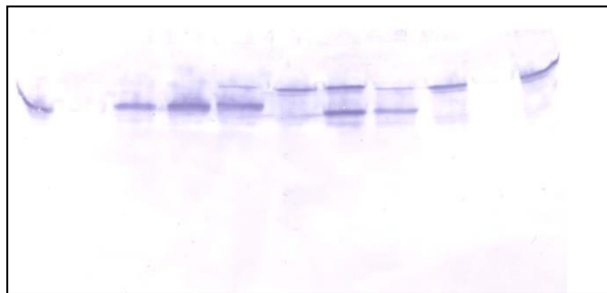
## 1. Polyakrylamidová gelová elektroforéza se specifickou detekcí :

**Detekce proteinu zájmu b hem jeho purifikace**

ÉPomocí protilátek

**Sledování biologické aktivity proteinu b hem purifikace**

ÉU enzym nap . barvení v gelu pomocí chromogenních substrát (nebo stanovení specifických konstant v komplexních vzorcích )



SDS PAGE s následných  
westernovým p enosem



nativní PAGE, zymogram



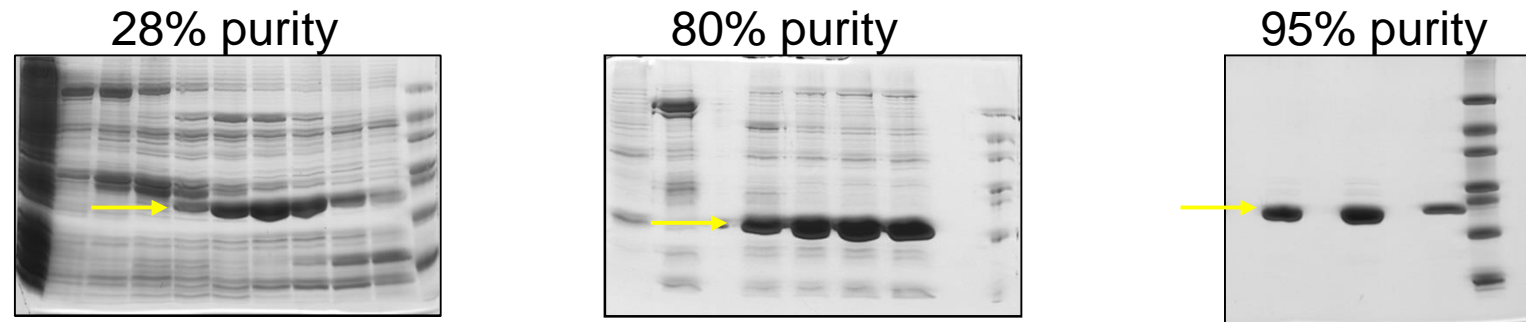
## 2. *Jak???*

## Jak budeme protein analyzovat?

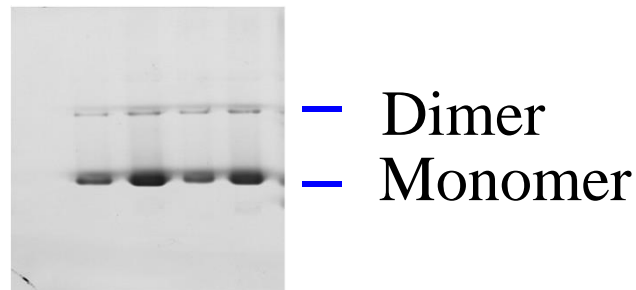
### 2. Polyakrylamidová gelová elektroforéza s nespecifickou detekcí

#### Sledování istoty a homogenity purifikovaného proteinu

ÉBarvení pomocí Coommasie blue, st íbra,...



#### Nativní PAGE



**Co???**

## **Jaké vlastnosti má protein ?**

Informace o proteinu zájmu a podobných proteinech z databází nebo z pilotních experimentů :

Velikost proteinu (SDS PAGE, gelová filtrace nebo analytická centrifugace)

Izoelektrický bod (izoelektrická fokusace)

Stabilita (pH, teplota, přítomnost solí, proteas, aditiv zajišťujících rozpustnost proteinu)

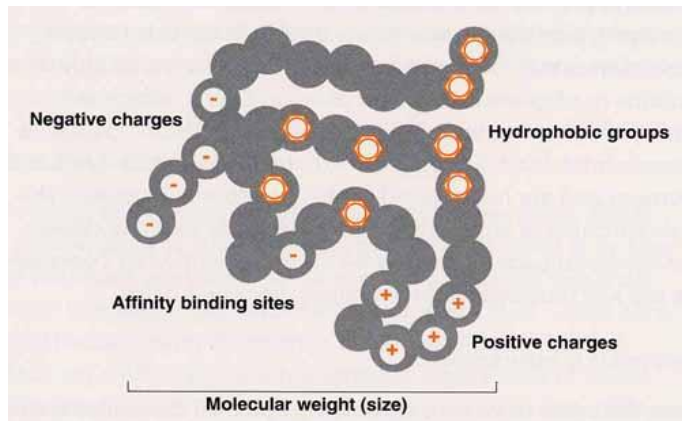
**2D a nativní PAGE**

Komplexita vzorku, vlastnosti proteinu zájmu a i ostatních kontaminujících proteinů

Informace o proteinu zájmu a o podobných proteinech z literatury:

Strategie purifikace (metody, pufrů, stabilita proteinu, ...)

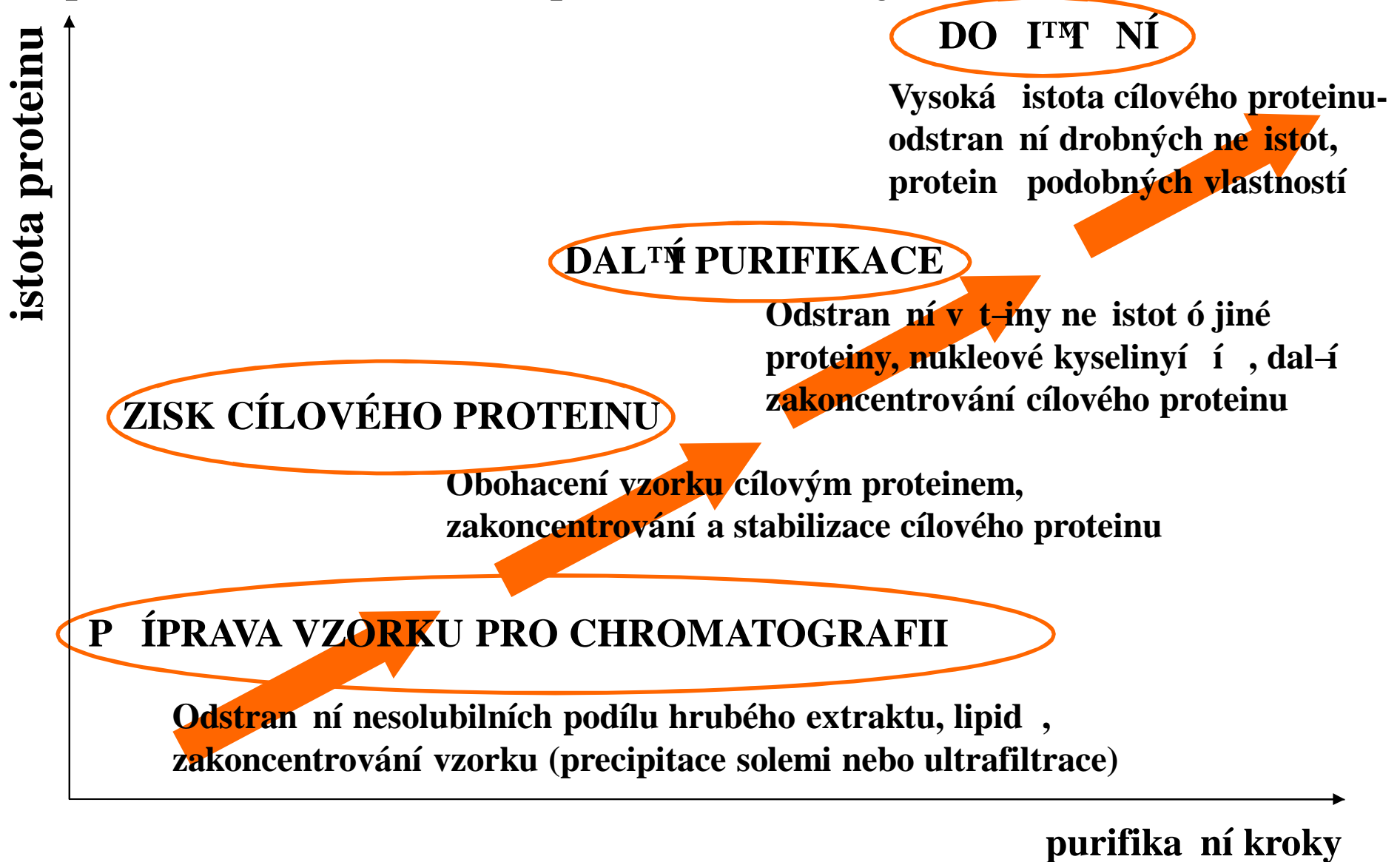
# Vlastnosti/purifikační metody



Rozpustnost	precipitace např. síranem amonným, nízké/vysoké pH
Stabilita	teplotní precipitace
Velikost	gelová filtrace (gelová permeační chromatografie)
pI (povrchový náboj)	iontovým výměnnou chromatografie
Hydrofobicita	reverzní fázová a hydrofóbní chromatografie
Specifická vazba	afinitní chromatografie
Posttranslační modifikace	afinitní chromatografie

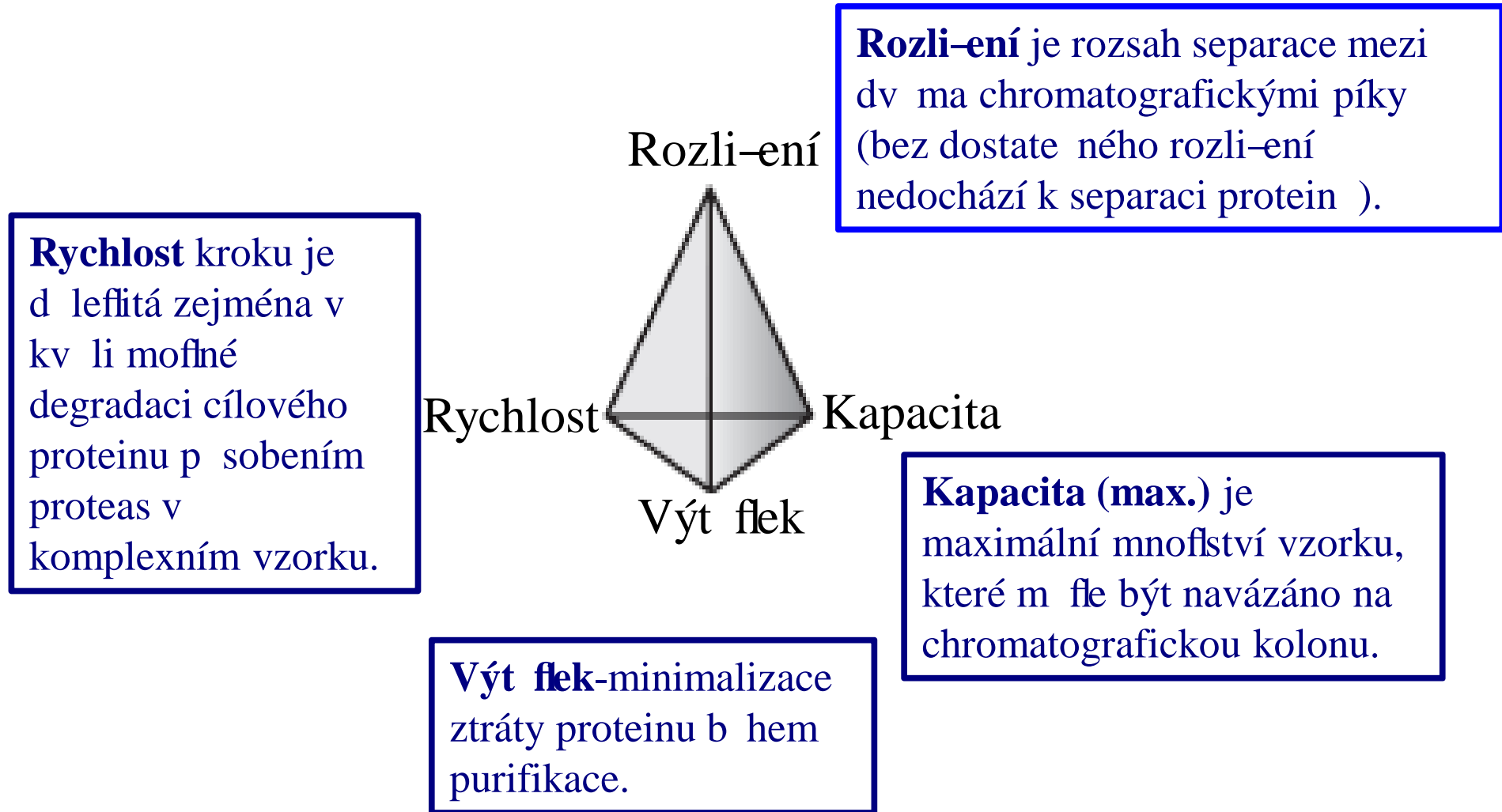
# Kolik purifikačních kroků je potřeba?

Příprava vzorku a tří fázová purifikační strategie



# Logická kombinace purifikačních kroků

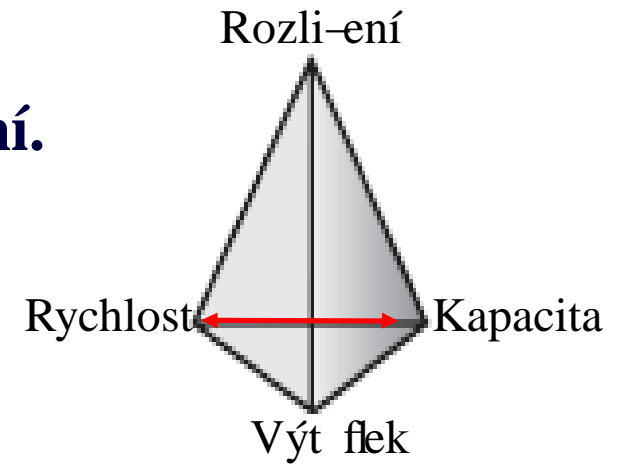
Každá separační technika je vyznačena rovnováhou mezi **tymi parametry**.



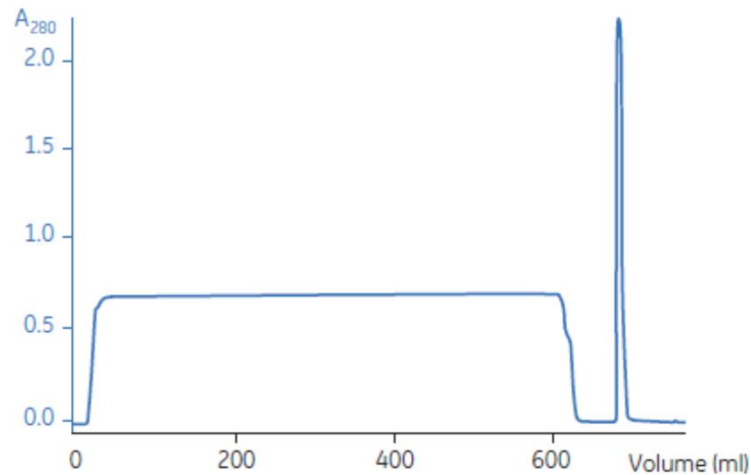
# Získ cílového proteinu z proteinového extraktu

**Cíl: rychlá izolace, stabilizace a zakoncentrování.**

**Purifikační techniky: afinitní chromatografie  
iontomní chromatografie  
hydrofóbní chromatografie**



*Column:* rProtein A Sepharose Fast Flow, XK16/20, bed height 4.8 cm (9.6 ml)  
*Sample:* 600 ml clarified cell culture containing 87.6 mg of IgG<sub>2a</sub>  
*Starting buffer:* 20 mM sodium phosphate, pH 7.0  
*Elution buffer:* 20 mM sodium citrate, pH 4.0  
*Flow rate:* 5 ml/min (150 cm/h)

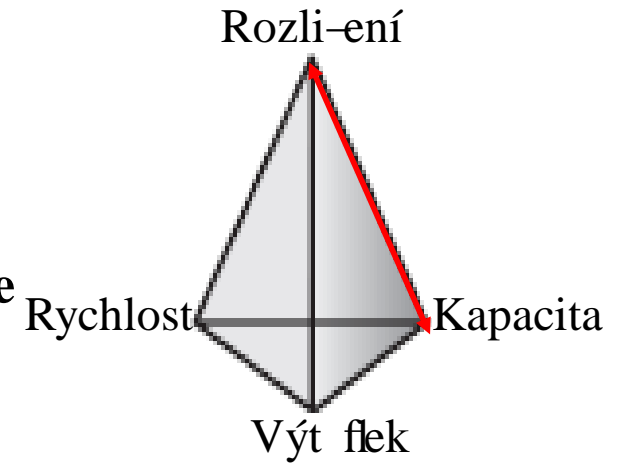


**Fig 4.5.** Example of capture step: Purification of IgG<sub>2a</sub> from clarified cell culture.

# Další purifikace proteinu

**Cíl:** Purifikace a zakoncentrování.

**Purifikační techniky:** iontomníová chromatografie  
hydrofóbní chromatografie  
gelová filtrace  
afinitní purifikace



Column: XK 16/20 Butyl Sepharose 4 Fast Flow  
Sample: 5 ml of partially purified Annexin V expressed in *E. coli*  
Buffer A: 20 mM Sodium phosphate, pH 7.0, 1 M  $(\text{NH}_4)_2\text{SO}_4$   
Buffer B: 20 mM Sodium phosphate, pH 7.0  
Flow rate: 100 cm/h  
Gradient: 0 to 50% B, 20 column volumes

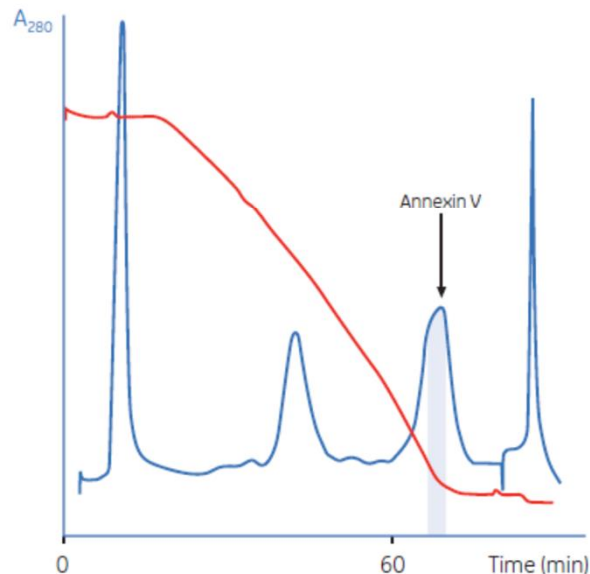


Fig 4.7. Example of an intermediate purification step: Purification of recombinant Annexin V by HIC.

# Do i-t ní proteinu a úprava podmínek pro jeho skladování (pH, soli, additiva)

**Cíl:** Produkt o požadované vysoké čistotě.

**Purifikační techniky:** gelová filtrace  
reverzní fázová chromatografie  
afinitní purifikace

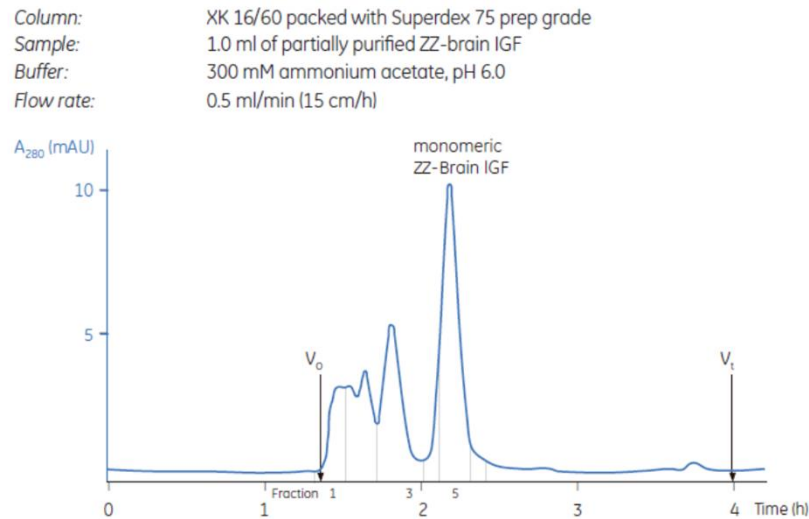
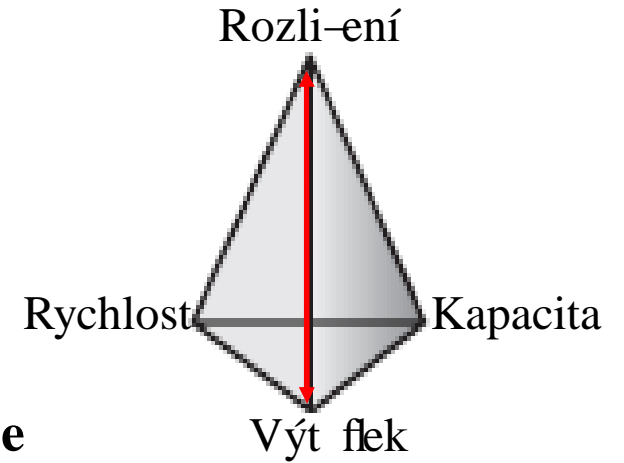


Fig 4.9. Example of polishing step: removal of dimers and multimers by GF.

Column: Mono S™ 5/50 GL  
Sample: 14.5 ml of partially purified and desalted transposase TniA  
Binding buffer: 20 mM MES pH 6.5, 1 mM EDTA, 2 mM MgCl<sub>2</sub>, 1 mM DTT  
Elution buffer: 20 mM MES pH 6.5, 1 mM EDTA, 2 mM MgCl<sub>2</sub>, 1 mM DTT, 1 M NaCl  
Flow rate: 1 ml/min  
Gradient: 0%–100% elution buffer, 20 CV

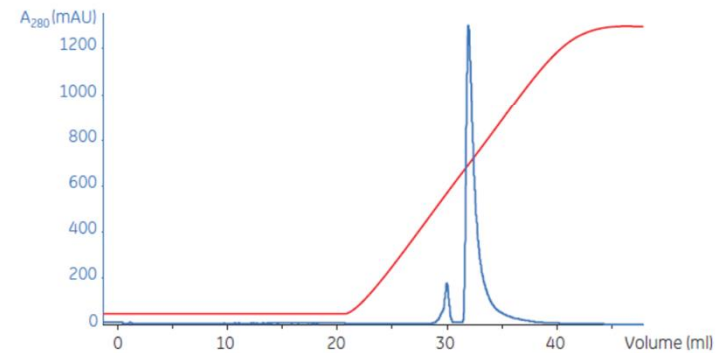


Fig 4.10. Example of polishing: removal of trace contaminants by high-resolution CIEC. Purification of the transposase TniA.



# Základní zásady pro pořadí purifikačních kroků

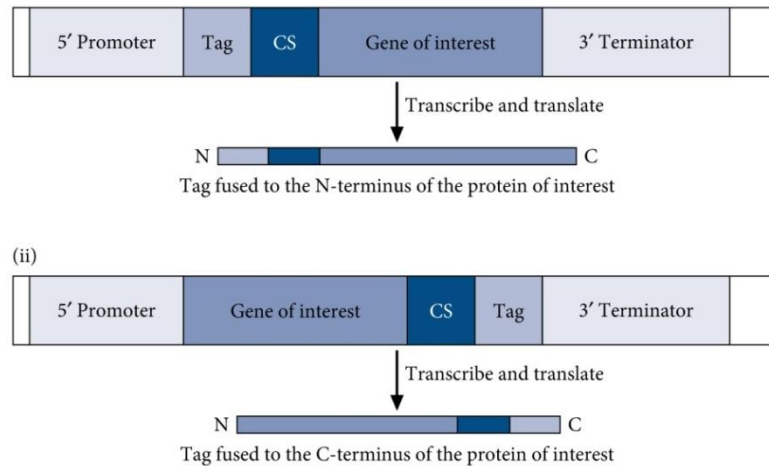
- É Na začátek zvolit metody s vysokou kapacitou a malým výtěkem a rozlišením → velké množství levného vstupního materiálu.
- É Později metody s vysokým rozlišením a výtěkem, kapacita méně významná → ve vzorku již investovaná práce, množství proteinu je menší.
- É Pokud možno zvolit metody za sebou racionálně, bez nutnosti mezikroků, jako je výměna pufru mezi jednotlivými separačními technikami  
příklad: po precipitaci síranem amonným nebo iontominióvé chromatografii (protein je eluován z kolony za vysokých koncentrací soli) zvolit hydrofóbní chromatografii (vzorek dávkován na kolonu ve vysoké koncentraci soli)
- É Jednotlivé separační metody pokud možno neopakovat.
- É Čím méně kroků, tím větší výtěčnost proteinu.

# Fúzní proteiny

Translační fúze sekvencí kódujících rekombinantní protein a tag (kotvu).

Tag : a) krátký peptid [p . (His)<sub>n</sub>, (Asp)<sub>n</sub>, (Arg)<sub>n</sub> ... ]

b) pirozený oligopeptid [p . MBP, GST, thioredoxin í ]



Úspěšná purifikace (uniformita purifikace) rekombinantního proteinu

Úspěšná výroba

Úspěšná rozpustnost

Úspěšná detekce

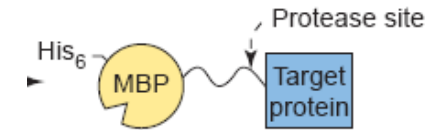
Úspěšná sekrece

Úspěšného partnera lze obvykle selektivně odtrhnout.

<b>Fúzní partner</b>	<b>Velikost</b>	<b>Umíst ní</b>	<b>Využití</b>
<b>His-tag</b>	6, 8, or 10 aa	N-, C-, internal	purifikace
<b>thioredoxin</b>	109 aa (11.7 kDa)	N-,C-	purifikace, zvý-ení solubility proteinu
<b>His-patch thioredoxin</b>	109 aa (11.7 kDa)	N-,C-	purifikace, zvý-ení solubility proteinu
<b>chloramfenikol acetyltransferasa</b>	24 kDa	N-	sekrece, purifikace, detekce
<b>avidin/streptavidin <i>Strep-tag</i></b>			purifikace, sekrece
<b>glutathion-S-transferasa-GST</b>	26 kDa	N-	purifikace
<b>maltosu vázající protein (MBP)</b>	40 kDa	N-, C-	purifikace, sekrece
<b>zelen fluoreskující protein (GFP)</b>	220 aa	N-, C-	detekce, purifikace
<b>polyasparagová kyselina</b>	5-16 aa	C-	purifikace
<b>ompT /ompA</b>	22 aa /21 aa	N-	sekrece

# Odstranění fúzní kotvy (tagu) a proteolytické –t pení

Jakákoliv kotva může vadit při funkčních a strukturních studiích rekombinantního proteinu.



## pRSET B Multiple Cloning Site

```

21  T7 promoter
    AATACGACTC ACTATAGGGA GACCACAACG GTTCCCTCT AGAAATAATT TTGTTAACT TTAAGAAGGA
                                     RBS

91  Polyhistidine (6xHis) region
    GATATACAT ATG CGG GGT TCT CAT CAT CAT CAT CAT CAT GGT ATG GCT AGC ATG ACT
    Met Arg Gly Ser His His His His His His Gly Met Ala Ser Met Thr

148 T7 gene 10 leader
    GGT GGA CAG CAA ATG GGT CGG GAT CTG TAC GAC GAT GAC GAT AAG GAT CCG AGC TCG
    Gly Gly Gln Gln Met Gly Arg Asp Leu Tyr Asp Asp Asp Asp Lys Asp Pro Ser Ser
    EK recognition site EK cleavage site

    Bgl II Pst I Pvu II Kpn I Nco I EcoR I BstB I Hind III
205 AGA TCT GCA GCT GGT ACC ATG GAA TTC GAA GCT TGA TCCGGCTGCT AACAAAGCCC
    Arg Ser Ala Ala Gly Thr Met Glu Phe Glu Ala ***

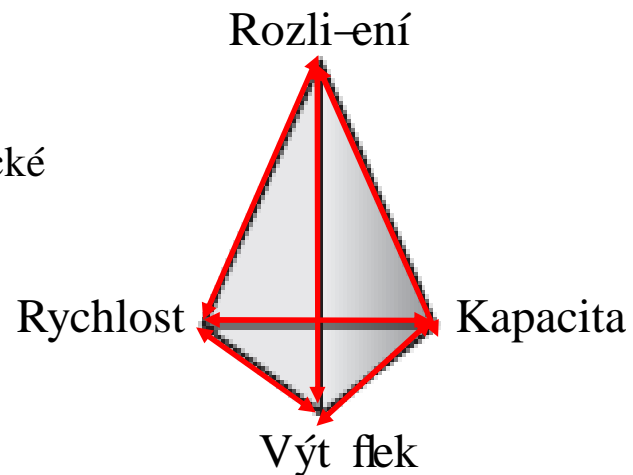
261 T7 reverse priming site
    GAAAGGAAGC TGAGTTGGCT GCTGCCACCG CTGAGCAATA ACTAGCATAA
  
```

Enzyme	Cleavage site
Enterokinase	DDDDK*
Factor Xa	IDGR*
Thrombin	LVPR*GS
PreScission	LEVLFQ*GP
TEV protease	EQLYFQ*G
3C protease	ETLFQ*GP
Sortase A	LPET*G
Granzyme B	D*X, N*X, M*N, S*X

## Fúzní kotvy (tagy) využívané se k purifikaci

Kotva (tag)	Typ chromatografie	Princip separa ní techniky
poly [His]	afinitní	vazba na kov
IgG vazná doména	afinitní	vazba na protilátku
Poly [Asp]	iontom ni ová	vazba na anion vázající matrici
Poly [Phe]	hydrofóbní	vazba na hydrofóbní matrici
<i>Strep</i> -tag	afinitní	vazba na streptavidin
Poly [Arg]	iontom ni ová	vazba na kation vázající matrici

Separa ní techniky charakteristické rovnováhou v-ech parametr .



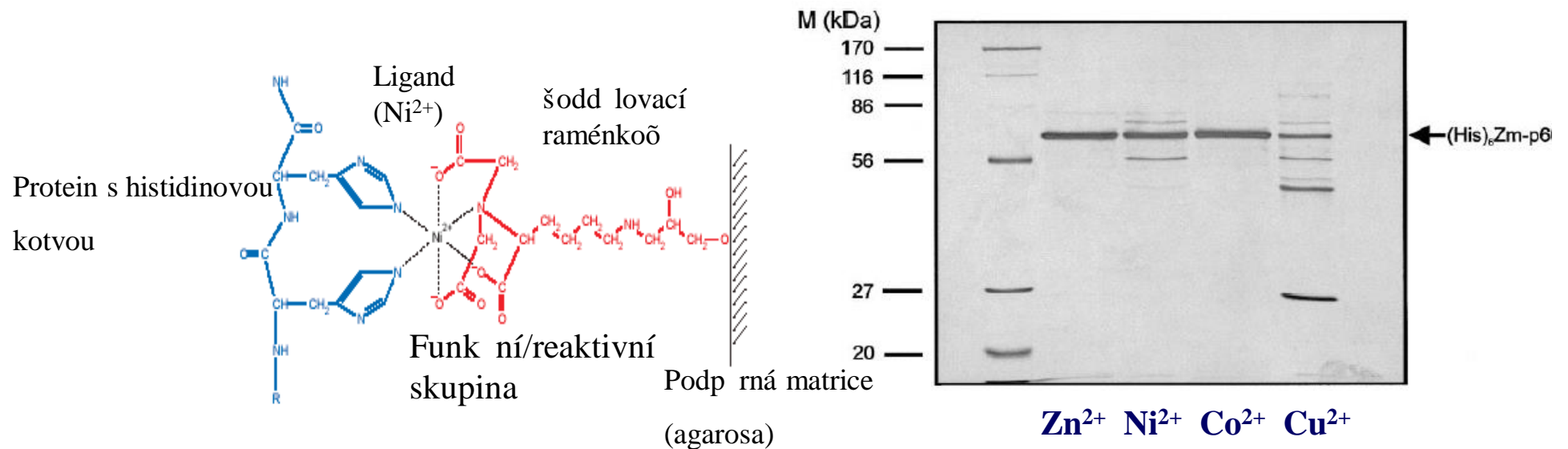
# Metalochelata ní afinitní chromatografie

ÉR.1975- uve ejnil Porath a kol. metodu frakcionace sérových protein

ÉKonstrukce um lých oligohistidinových domén fúzovaných k N- nebo C- konci proteinu metodami mol. biologie

ÉNyní jeden ze základních purifika ních postup rekombinantních protein

ÉInterakce proteinu s matricí je zprost edkovaná neobsazenými d-orbitaly iont p echodných kov , které váflou volné elektronové páry p eváfln z dusíkového atomu imidazolových skupin histidinových zbytk v proteinu.



Síla vazby: Cu<sup>2+</sup> > Ni<sup>2+</sup> > Zn<sup>2+</sup> ~ Co<sup>2+</sup>

# Metalochelata ní afinitní chromatografie

## *Purifikace za nativních podmínek*

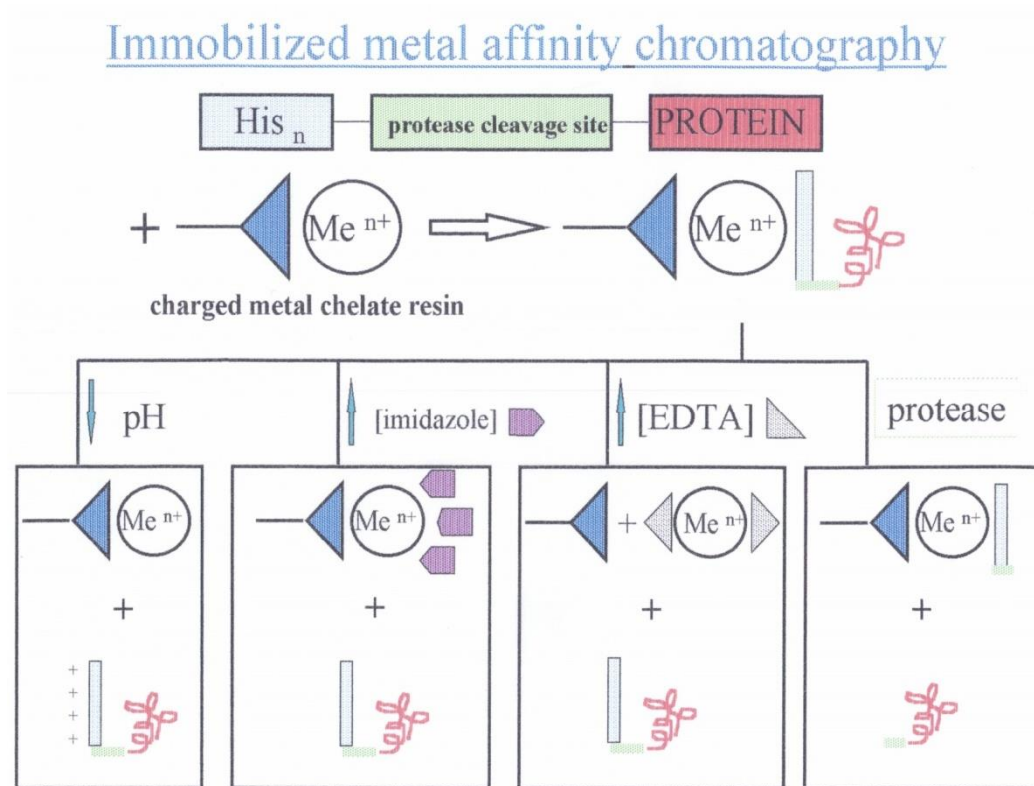
### Obecně lze navrhnout:

Pufry o pH 7-8 pro vazbu rek. proteinu s kovovým iontem.

Pufry s vysokou koncentrací solí (nap. 0,5-1 mol/l NaCl).

Nízká koncentrace imidazolu nebo snížení pH pro odstranění balastních proteinů.

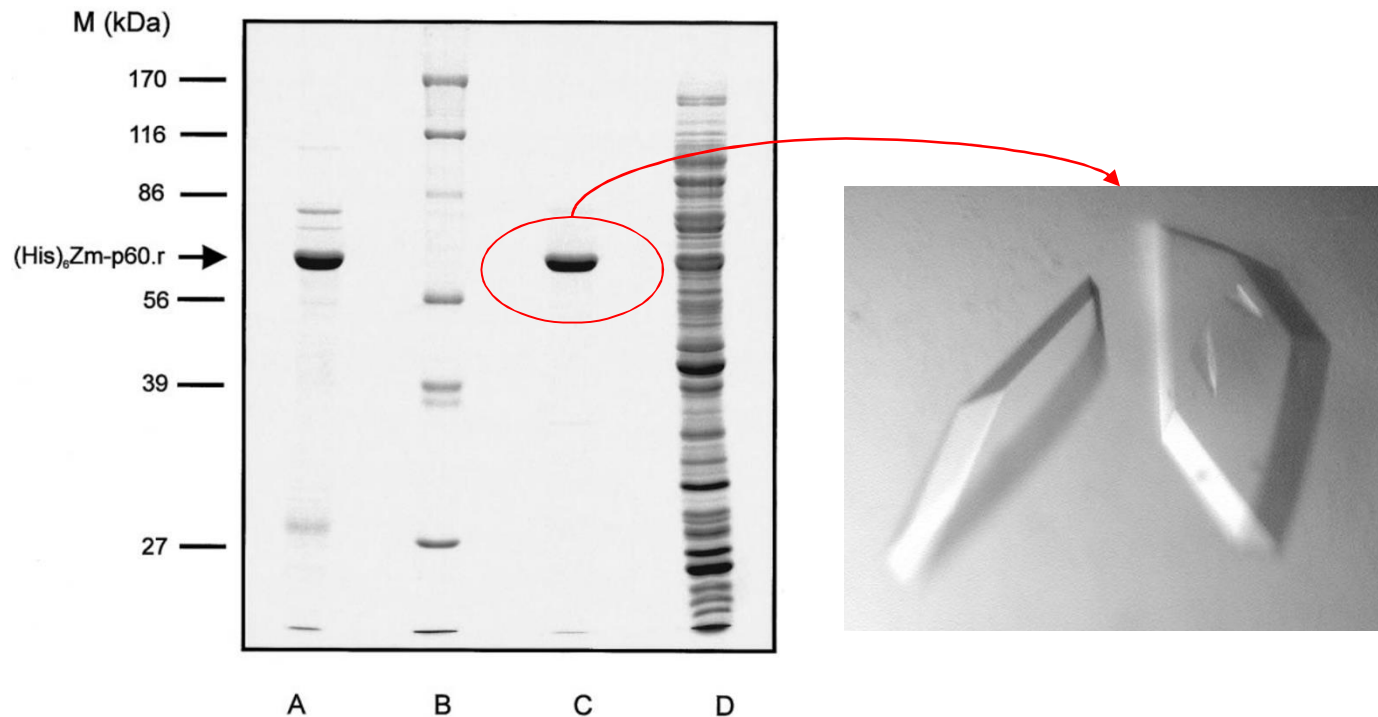
Eluce použitím gradientu imidazolu (0-1 mol/l), výrazným snížením pH nebo vyfilitím EDTA.



# His-tagged protein and IMAC under native conditions

## One-step purification of maize $\beta$ -glucosidase

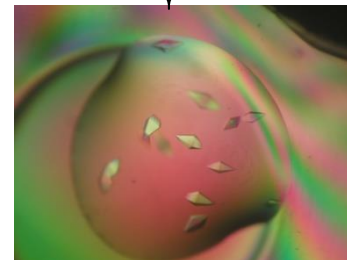
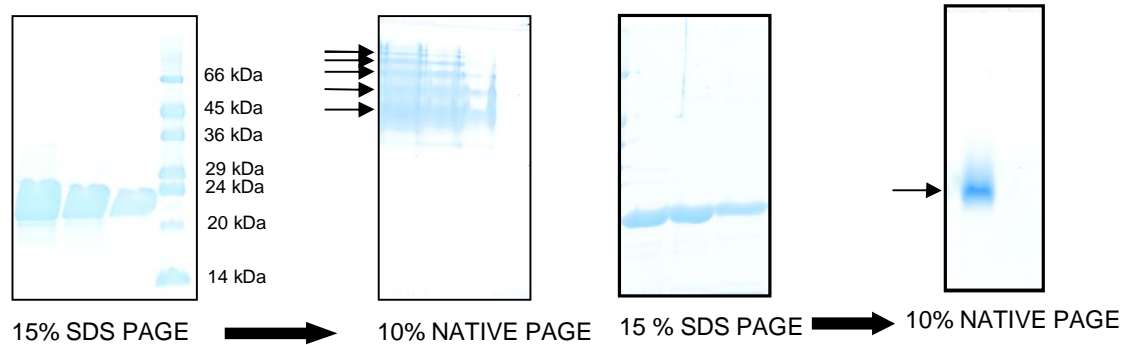
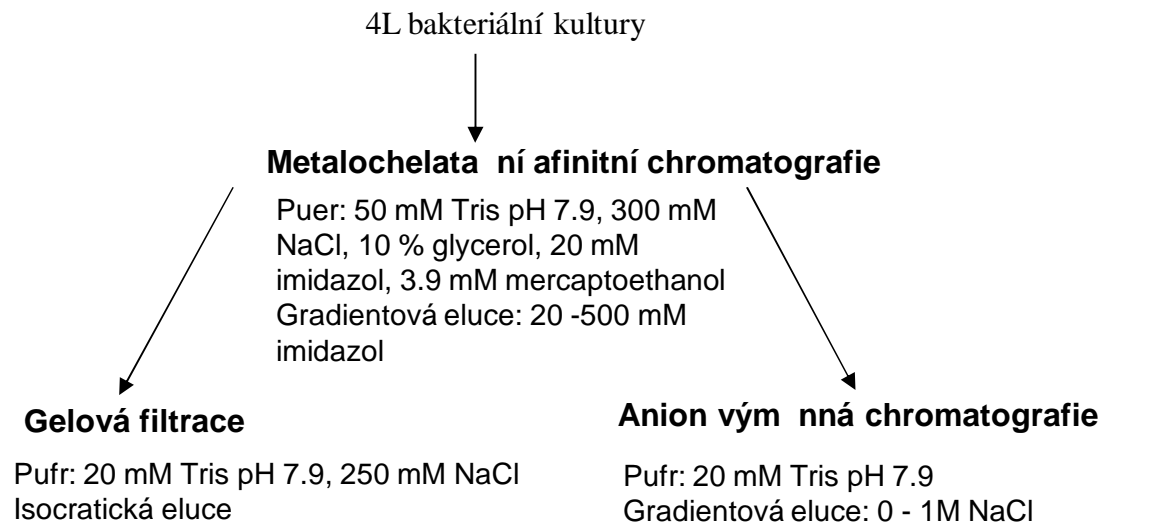
- Perfusion matrix: POROS MC/M
- Functional group: iminodiacetate, metal ion  $Zn^{2+}$
- Removing contaminated proteins: linear gradient of imidazole (0 to 50 mM) and pH (pH 7 to 6.1)
- Protein elution: 0.1 M EDTA
- 80% recovery, 95 fold purification
- Common production and isolation of wild type protein and soluble mutant form for enzymatic measurements and crystallization.



(Zouhar et al., 1999)



# Purifikace proteinu AHP2 (Arabidopsis histidin phosphotransfer protein 2)



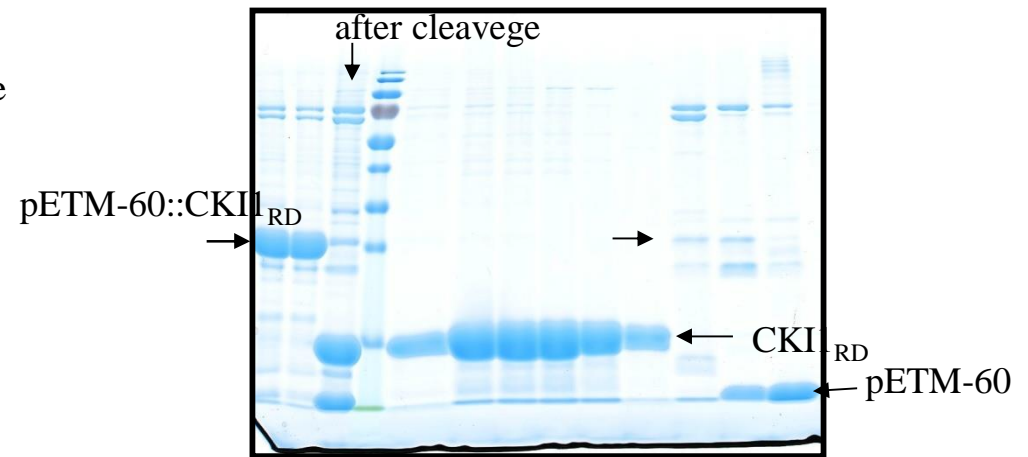
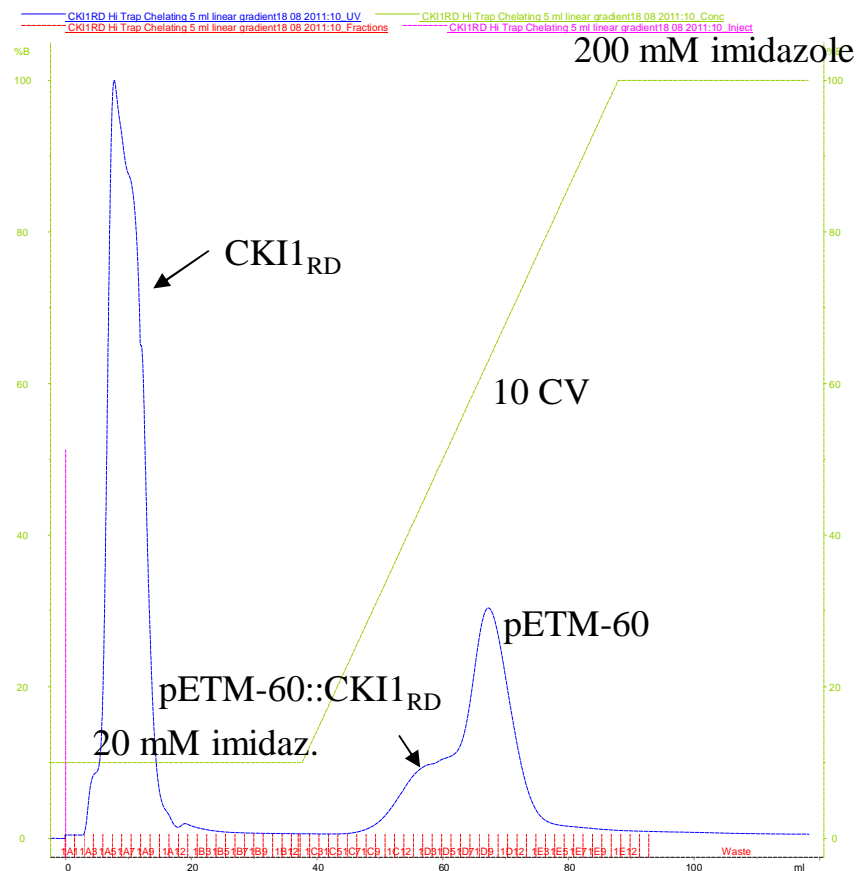
# His-tagged protein and IMAC under native conditions

## Four-step purification of *Arabidopsis* CKI<sub>RD</sub>

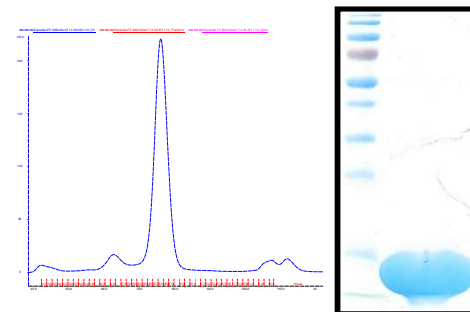
1. Affinity purification (MCAC)
2. Tag removal (TEV protease)
3. Affinity purification (MCAC)
4. Size exclusion chromatography



### 3. Affinity purification after TEV cleavage



### 4. Size-exclusion chromatography



1 L ~10-20 mg for TB and M9

# Doporu ená literatura

Makrides SV (1996) **Strategies for Achieving High-Level Expression of Genes in *Escherichia coli***. Microb.Review 60, (512-538).

Simpson RJ; Adams PD; Golemis E

**Basic methods in protein purification and analysis: a laboratory manual**,  
Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, 2009,  
436 s., ISBN 978-0-87969-868-3

