Sylabus – Přednášky a semináře C5855 a C5856, podzim 2024

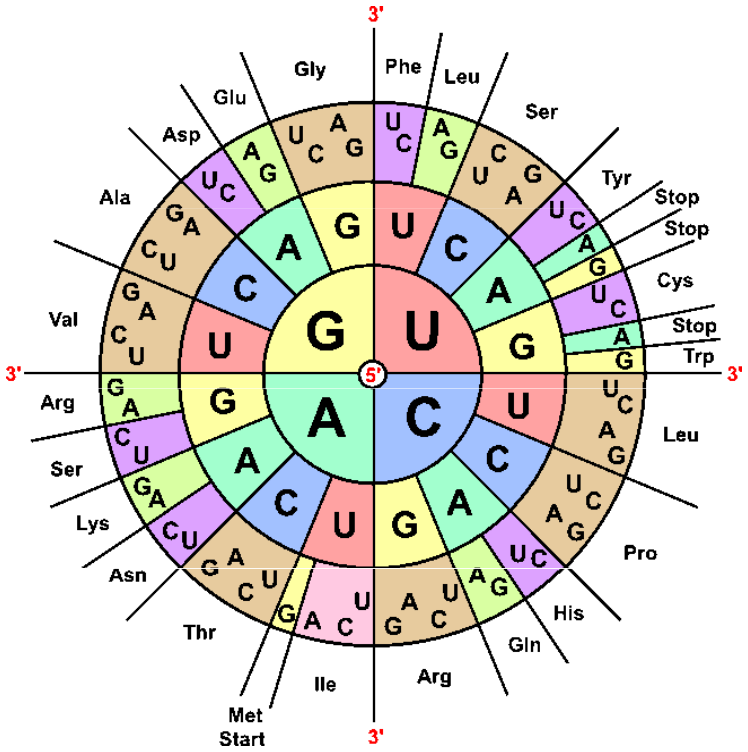
Datum	Přednáška	Seminář
24.9.	Úvod do studia, Koncepce přípravy rekombinantních proteinů, Základní analytické pojmy	x
1.10.	Separační a Chromatografické metody, Ostatní separační metody: Fokusace, Ultracentrifugace, Dialýza	Úloha 1_TLC
8.10.	Spektroskopické techniky I : UV-Vis, Fluorescence, FRET	x
15.10.	Spektroskopické techniky II: CD, vibrační spektroskopie: IC, Raman, X-Ray, QD	Úloha 2_CD
22.10.	NMR a EPR	x
29.10.	Optická a elektronová mikroskopie (SEM,TEM), AFM, Fluorescenční mikroskopie, Konfokální mikroskopie, Klasická mikroskopie	x
5.11.	MS, LC-MS, Aplikace MS	Úloha 3_MALDI-TOF MS
12.11.	Úvod do elektrochemie (pH, pKa, Nernstova rovnice, Voltametrie, Potenciometrie, Amperometrie, Impedanční a Pulzní Voltametrie + jejich aplikace)	x
19.11.	Elektroforéza - Gelová ELFO, Kapilární ELFO, Aplikace	Úloha 4_SDS-PAGE
26.11.	Bio-elektrochemie - Rozptyl, X-Ray, Biosenzory, Aplikace, SPR.	x
3.12.	Strukturní biochemie - Techniky pro určování 3D struktur proteinu (X-ray, NMR, cryoEM, simulované žíhání)	Exkurze 1, pavilon C4 (Laboratoř přípravy proteinu a NMR park)
10.12.	Biointerakce - Metody pro určování interakcí a oligomerních stavů - BioKalorimetrie (Termostabilita, ITC a DSC + alternativní interakce/hydrodynamické techniky (osmometrie, centrifugace, Svedberg).	Exkurze 2, pavilon C4 (ITC, DLS, nanoDSF...)
17.12.	Metody lékařské diagnostiky	Exkurze 3, pavilon C14 (MS, MALDI-TOF MS, ESI-Orbitrap, DROBECEK (MS))

Exprese a purifikace proteinů

Jozef Hritz

(jozef.hritz@ceitec.muni.cz)

Central Dogma of Molecular Biology



Codon usage in *E. coli* genes

	Codon	Amino Acid		%	Codon	Amino Acid		%	Codon	Amino Acid		%	Codon	Amino Acid		%	
U	UUU	Phe	F	0.51	UCU	Ser	S	0.19	UAU	Tyr	Y	0.53	UGU	Cys	C	0.43	U
	UUC	Phe	F	0.49	UCC	Ser	S	0.17	UAC	Tyr	Y	0.47	UGC	Cys	C	0.57	C
	UUA	Leu	L	0.11	UCA	Ser	S	0.12	UAA	STOP		0.62	UGA	STOP		0.30	A
	UUG	Leu	L	0.11	UCG	Ser	S	0.13	UAG	STOP		0.09	UGG	Trp	W	1.00	G
C	CUU	Leu	L	0.10	CCU	Pro	P	0.16	CAU	His	H	0.52	CGU	Arg	R	0.42	U
	CUC	Leu	L	0.10	CCC	Pro	P	0.10	CAC	His	H	0.48	CGC	Arg	R	0.37	C
	CUA	Leu	L	0.03	CCA	Pro	P	0.20	CAA	Gln	Q	0.31	CGA	Arg	R	0.05	A
	CUG	Leu	L	0.55	CCG	Pro	P	0.55	CAG	Gln	Q	0.69	CGG	Arg	R	0.08	G
A	AUU	Ile	I	0.47	ACU	Thr	T	0.21	AAU	Asn	N	0.39	AGU	Ser	S	0.13	U
	AUC	Ile	I	0.46	ACC	Thr	T	0.43	AAC	Asn	N	0.61	AGC	Ser	S	0.27	C
	AUA	Ile	I	0.07	ACA	Thr	T	0.30	AAA	Lys	K	0.76	AGA	Arg	R	0.04	A
	AUG	Met	M	1.00	ACG	Thr	T	0.23	AAG	Lys	K	0.24	AGG	Arg	R	0.03	G
G	GUU	Val	V	0.29	GCU	Ala	A	0.19	GAU	Asp	D	0.59	GGU	Gly	G	0.38	U
	GUC	Val	V	0.20	GCC	Ala	A	0.25	GAC	Asp	D	0.41	GGC	Gly	G	0.40	C
	GUA	Val	V	0.17	GCA	Ala	A	0.22	GAA	Glu	E	0.70	GGA	Gly	G	0.09	A
	GUG	Val	V	0.34	GCG	Ala	A	0.34	GAG	Glu	E	0.30	GGG	Gly	G	0.13	G
	U				C				A				G				

Codon		Standard code (Translation table 1)		Name
DNA	RNA			
TAG	UAG	STOP = Ter	(*)	"amber"
TAA	UAA	STOP = Ter	(*)	"ochre"
TGA	UGA	STOP = Ter	(*)	"opal" (or "umber")

```
UUU F 0.57   UCU S 0.11   UAU Y 0.53   UGU C 0.42
UUC F 0.43   UCC S 0.11   UAC Y 0.47   UGC C 0.58
UUA L 0.15   UCA S 0.15   UAA * 0.64   UGA * 0.36
UUG L 0.12   UCG S 0.16   UAG * 0.00   UGG W 1.00

CUU L 0.12   CCU P 0.17   CAU H 0.55   CGU R 0.36
CUC L 0.10   CCC P 0.13   CAC H 0.45   CGC R 0.44
CUA L 0.05   CCA P 0.14   CAA Q 0.30   CGA R 0.07
CUG L 0.46   CCG P 0.55   CAG Q 0.70   CGG R 0.07

AUU I 0.58   ACU T 0.16   AAU N 0.47   AGU S 0.14
AUC I 0.35   ACC T 0.47   AAC N 0.53   AGC S 0.33
AUA I 0.07   ACA T 0.13   AAA K 0.73   AGA R 0.02
AUG M 1.00   ACG T 0.24   AAG K 0.27   AGG R 0.03

GUU V 0.25   GCU A 0.11   GAU D 0.65   GGU G 0.29
GUC V 0.18   GCC A 0.31   GAC D 0.35   GGC G 0.46
GUA V 0.17   GCA A 0.21   GAA E 0.70   GGA G 0.13
GUG V 0.40   GCG A 0.38   GAG E 0.30   GGG G 0.12

[Codon/a.a./fraction per codon per a.a.]
E. coli K12 data from the Codon Usage Database
```

Different codon usage

e.g. BL21-Codon plus-RIL

Codon Usage in E. coli & humans

Codon	Amino acid	Frequency of use in:	
		E. coli	Humans
GAG	Glutamic acid	0.30	0.59
GAA	Glutamic acid	0.70	0.41
CGG	Arginine	0.08	0.19
CGA	Arginine	0.05	0.10
CGU	Arginine	0.42	0.09
CGC	Arginine	0.37	0.19
AGG	Arginine	0.03	0.22
AGA	Arginine	0.04	0.21
CCG	Proline	0.55	0.11
CCA	Proline	0.20	0.27
CCU	Proline	0.16	0.29
CCC	Proline	0.10	0.33
UGA	Stop	0.30	0.61
UAG	Stop	0.09	0.17
UAA	Stop	0.62	0.23

Codon optimized sequence can be often achieved by synthetic gene synthesis - > particularly useful for the projects where the large expressions are needed, e.g. structural biology. Or isotopic labeling…

Often cleavage sites are considered

Thrombin (optimal pH ~8.0)

- Pro-Arg/Gly
- Pro-Lys/Leu
- Ala-Arg/Gly
- Gly-Lys/Ala
- Ile-Arg/Ser
- Leu-Arg/Ala
- Ile-Arg/Ile

TEV protease

- Glu-Asn-Leu-Tyr-Phe-Gln/Ser
- pH 5.5 –8.5

Phusion Tags

- a) short peptides [ex. (His)n, (Asp)n, (Arg)n ..
- b) protein domains, entire proteins

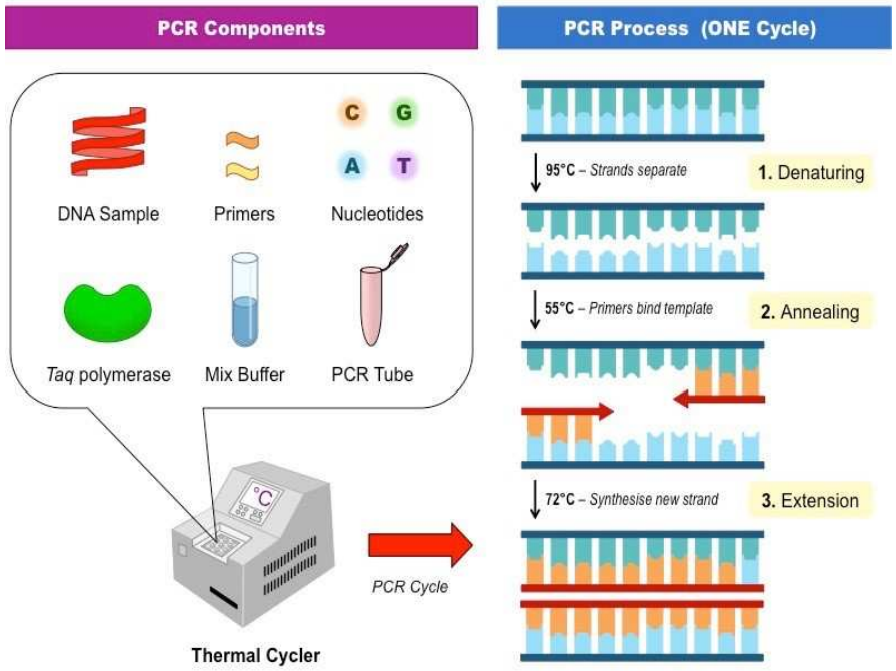
  [ex. MBP, GST, thioredoxin …].

Fusion partner (tag)	Size	Tag placement	Uses
His-tag	6, 8, or 10 aa	N- or C-terminus	Purification, detection
Thioredoxin	109 aa (11.7 kDa)	N- or C-terminus	Purification, solubility enhancement
Calmodulin-binding domain (CBD)	26 aa	N- or C-terminus	Purification
Avidin/streptavidin *Strep*-tag	8 aa	N- or C-terminus	Purification, secretion
Glutathione *S*-transferase (GST)	26 kDa	N-terminus	Purification, solubility enhancement
Maltose binding protein (MBP)	396 aa (40 kDa)	N- or C-terminus	Purification, solubility enhancement
Green fluorescent protein (GFP)	220 aa (27 kDa)	N- or C-terminus	Localization, detection, purification
Poly-Arg	5-16 aa	N- or C-terminus	Purification, solubility enhancement
N-utilization substance A (NusA)	495 aa (54.8 kDa)	N-terminus	Solubility enhancement

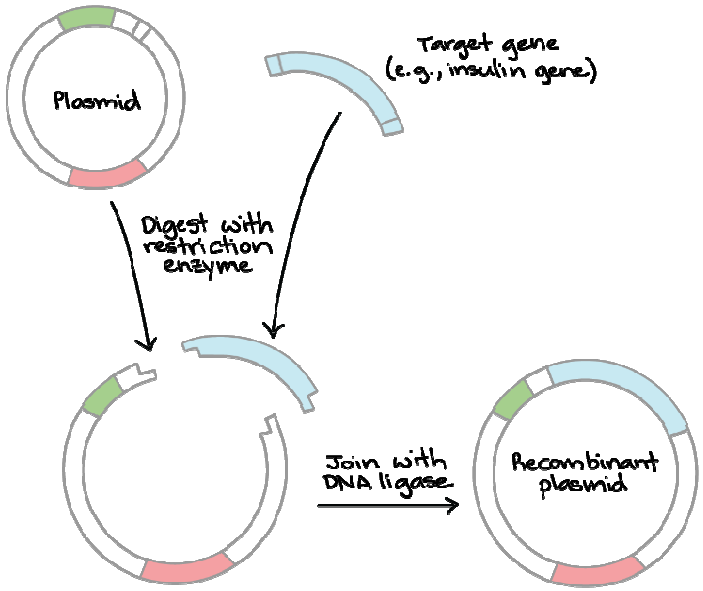
Purpose

- Increasing the yield of recombinant proteins – Fusion of the N-terminus of the target protein to the C-terminus of a highly expressed fusion partner results in high level expression of the target protein.
- Enhancing the solubility of recombinant proteins – Fusion of the N-terminus of the target protein to the C-terminus of a soluble fusion partner often improves the solubility of the target protein.
- Facilitating the purification of recombinant proteins – Simple purification schemes have been developed for proteins used at either terminus which bind specifically to affinity resins.

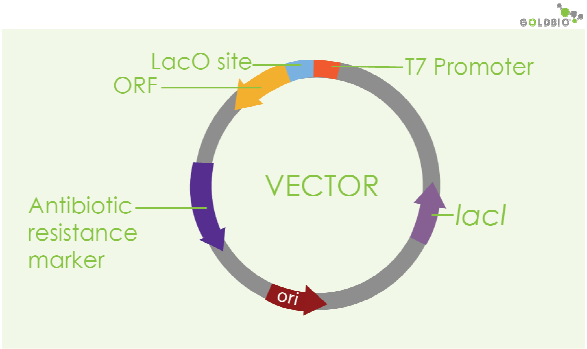
Once there is clear idea of DNA sequence

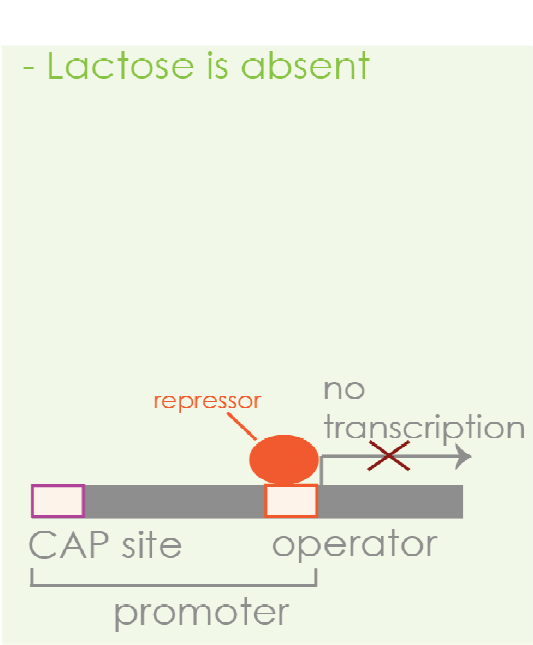


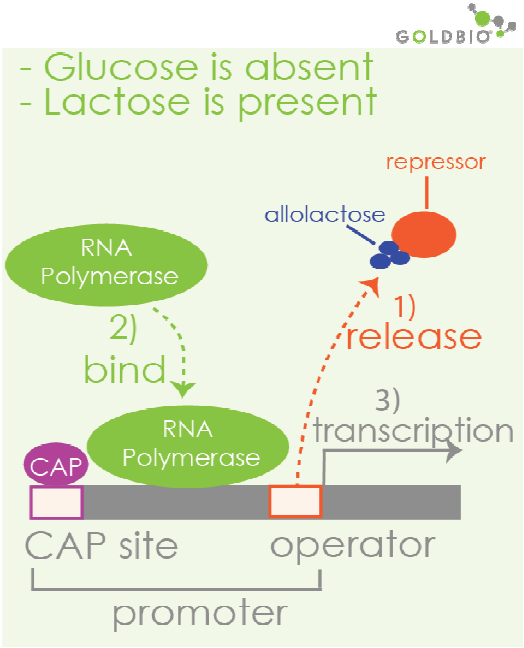
Gene amplification by PCR
insertion into the vector by e.g. restriction endonucleases

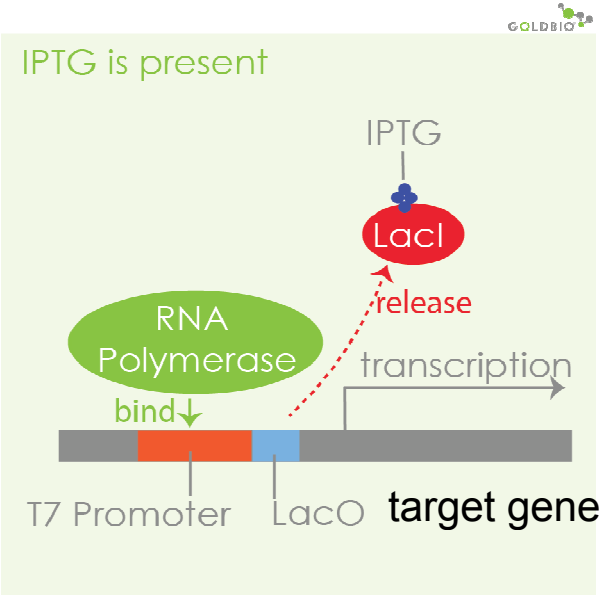


- circular E.coli plasmids (in addition to E.coli genome) – why antibiotic resistance is needed?
- Lac operator









pET plasmids

1.3. The primary structure of proteins

1.3.2. Information available from the amino acid sequence of a protein

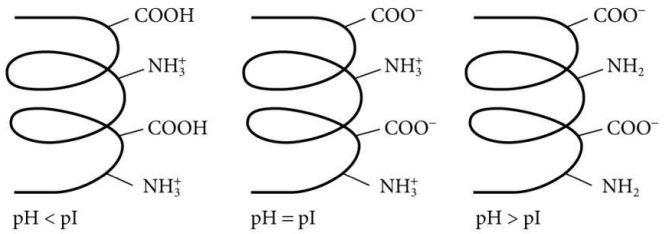
1.3.2.1. Exact molecular mass

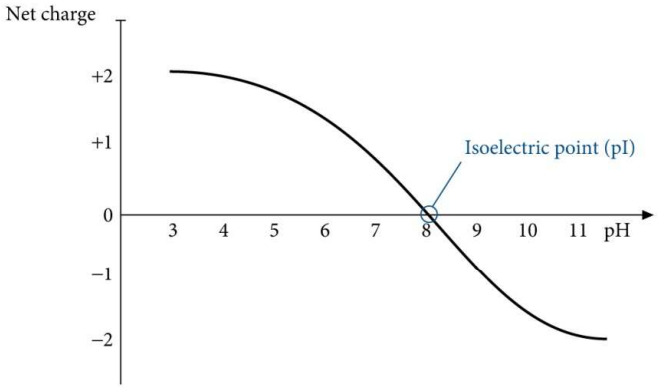
1.3.2.2. Isoelectric point

1.3.2.3. Absorption coefficient

1.3.2.4. Hydrofobicity

http://www.expasy.ch/tools/pi_tool.html





1.2. The amino acids

1.2.1. The variety of amino acids

1.2.2. Clasification of the amino acids in terms of polarity

Non-polar side chain

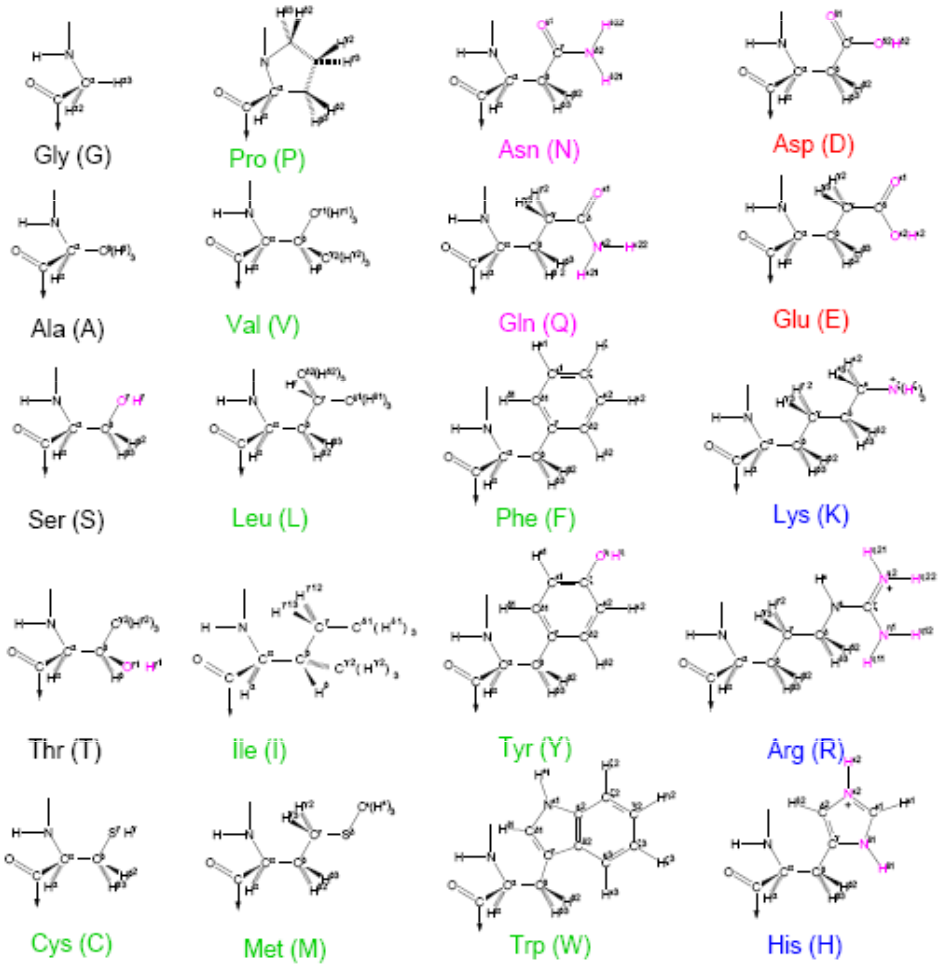
Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, Val

Polar, uncharged side chain

Asn, Cys, Gln, Ser, Thr, Tyr

Polar charged side chain

Arg, Asp, Glu, His, Lys

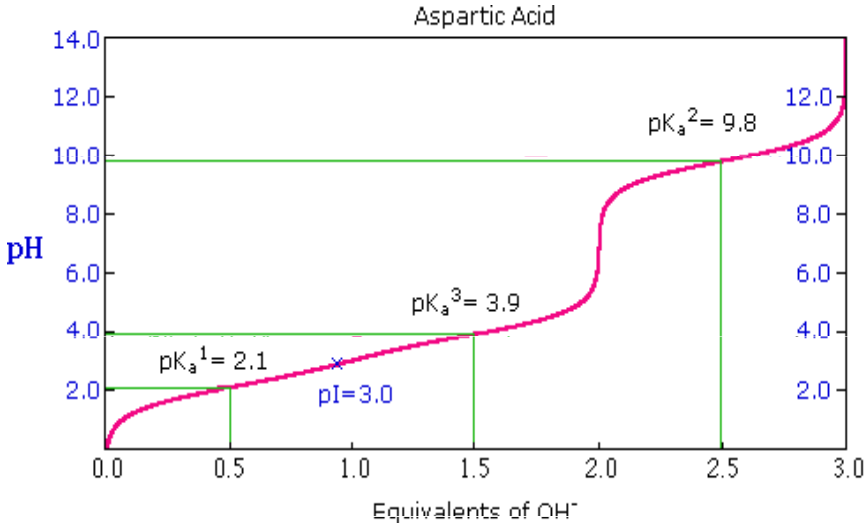


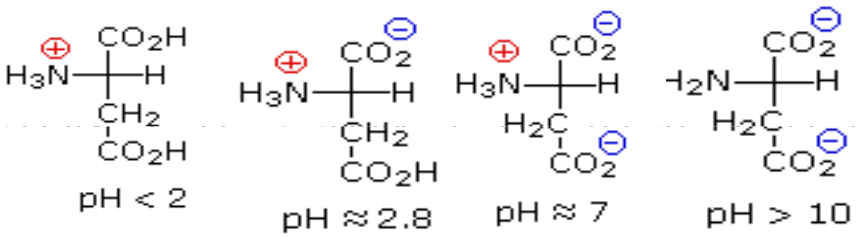
1.2. The amino acids

1.2.3. General properties of the amino acids

$$pH = \log_{10}\left[\frac{1}{a_{H^+}}\right] \cong \log_{10}\left[\frac{1}{[H^+]}\right]$$

1.2.3.2. Ionization

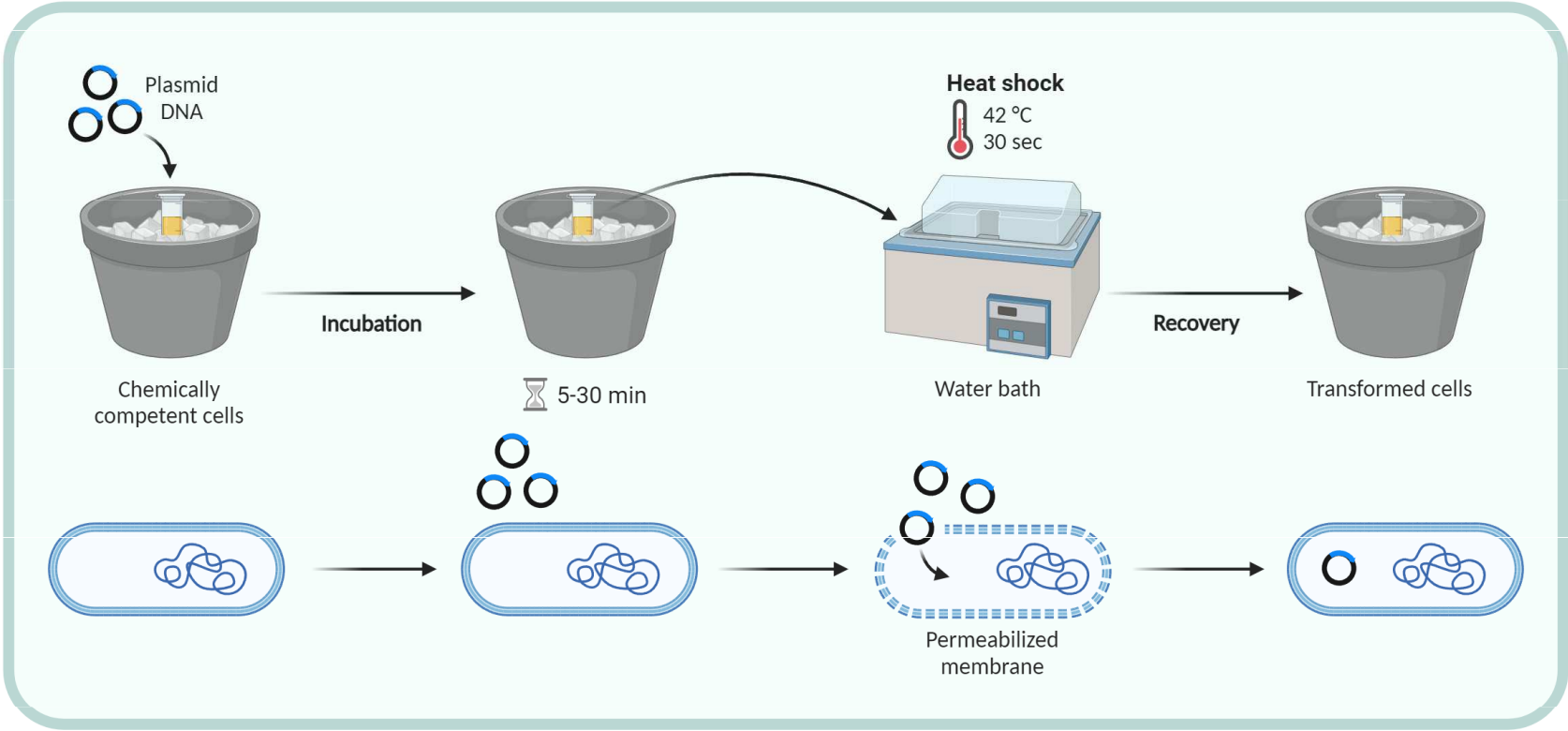




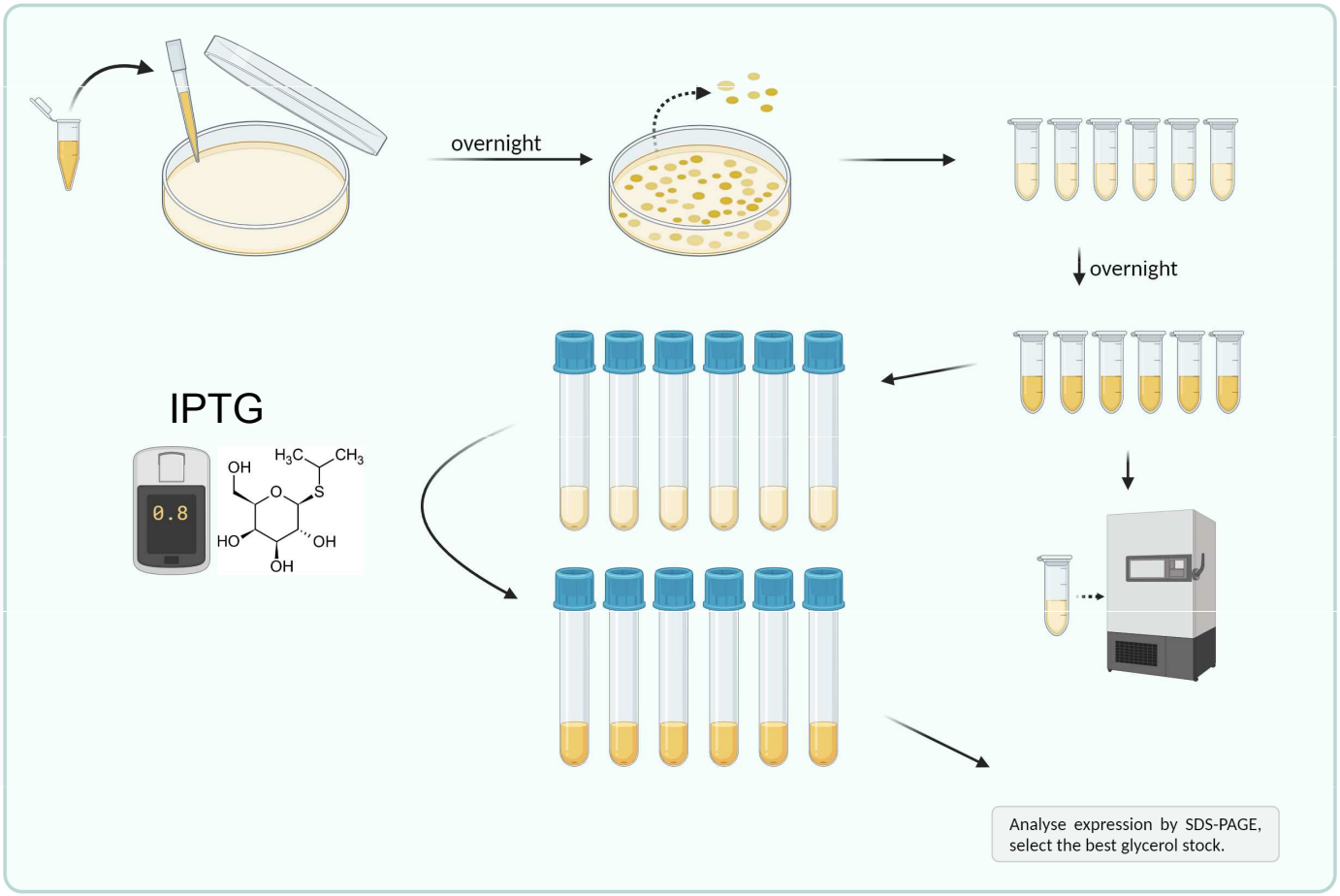
Amino Acid	Symbol	pK_1 (COOH)	pK_2 (NH2)	pK R Group
Glycine	Gly	2,4	9,8	
Alanine	Ala	2,4	9,9	
Valine	Val	2,2	9,7	
Leucine	Leu	2,3	9,7	
Isoleucine	Ile	2,3	9,8	
Serine	Ser	2,2	9,2	
Threonine	Thr	2,1	9,1	
Cysteine	Cys	1,9	10,8	8,3
Methionine	Met	2,1	9,3	
Aspartic Acid	Asp	2	9,9	3,9
Glutamic Acid	Glu	2,1	9,5	4,1
Asparagine	Asn	2,1	8,8	
Glutamine	Gln	2,2	9,1	
Arginine	Arg	1,8	9	12,5
Lysine	Lys	2,2	9,2	10,8
Histidine	His	1,8	9,2	6
Phenylalanine	Phe	2,2	9,2	
Tyrosine	Tyr	2,2	9,1	10,1
Tryptophan	Trp	2,4	9,4	
Proline	Pro	2	10,6	

http://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/proteins.ht

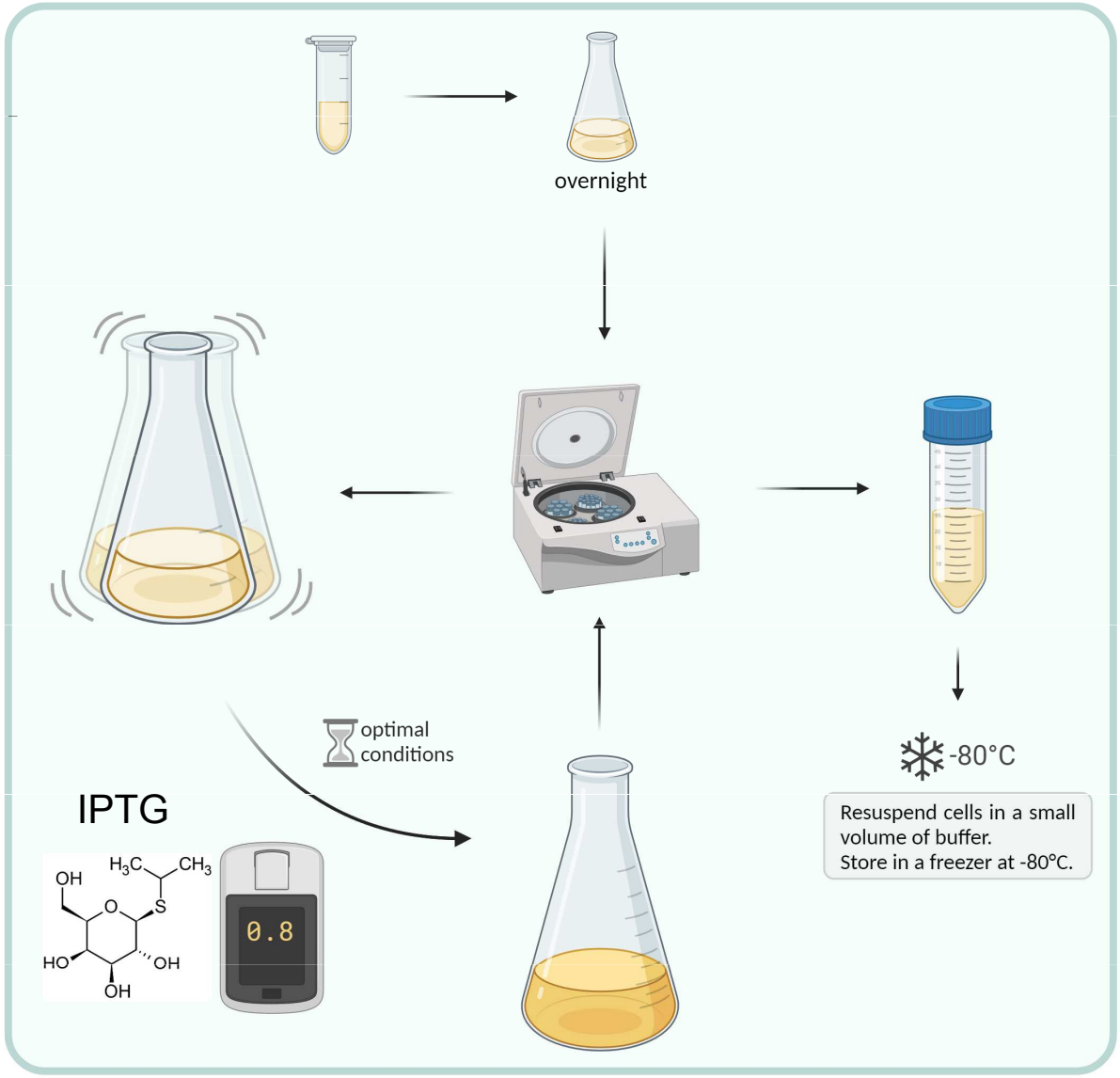
Transformace



Colony selection

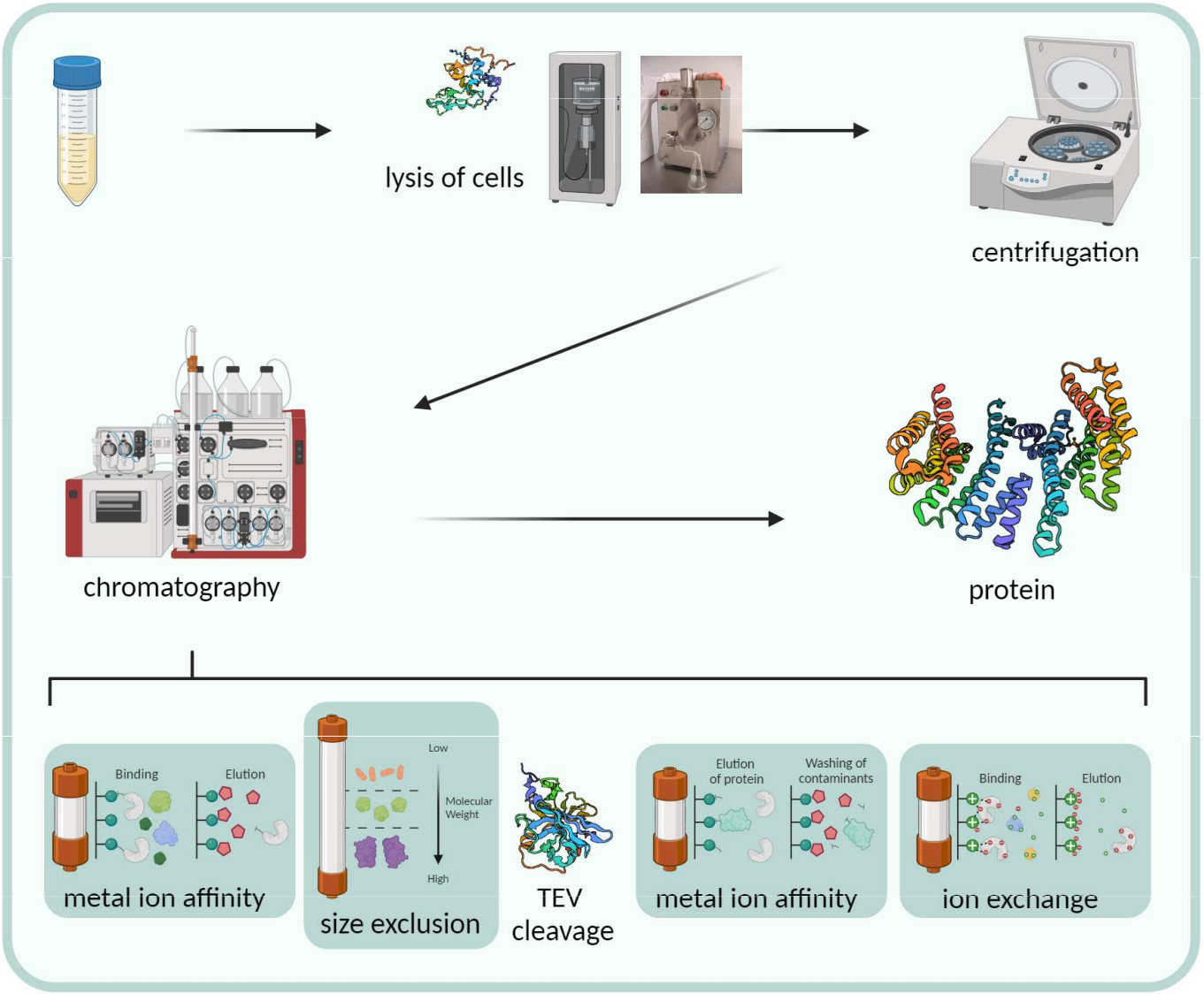


Expres

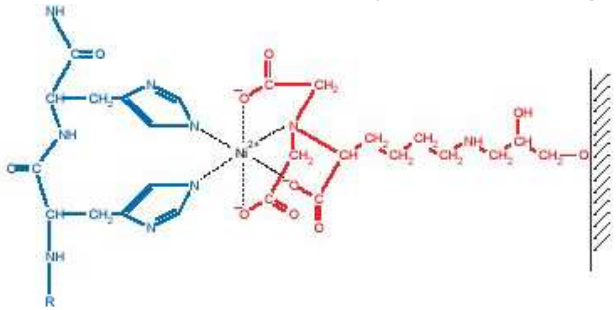


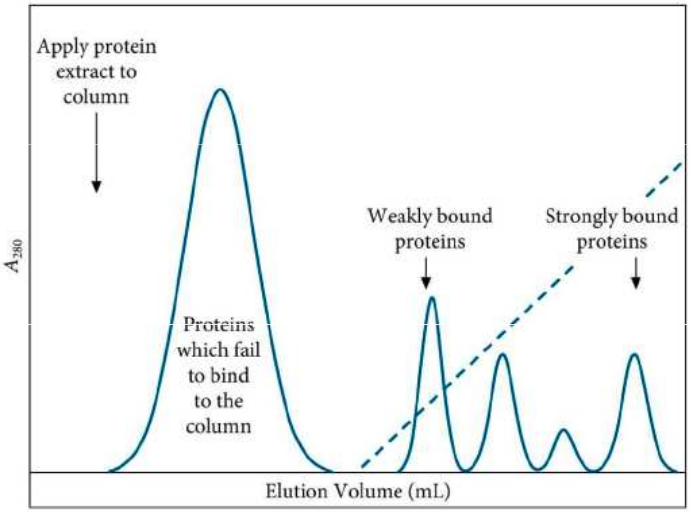
What kind of growing media?

Purifikace

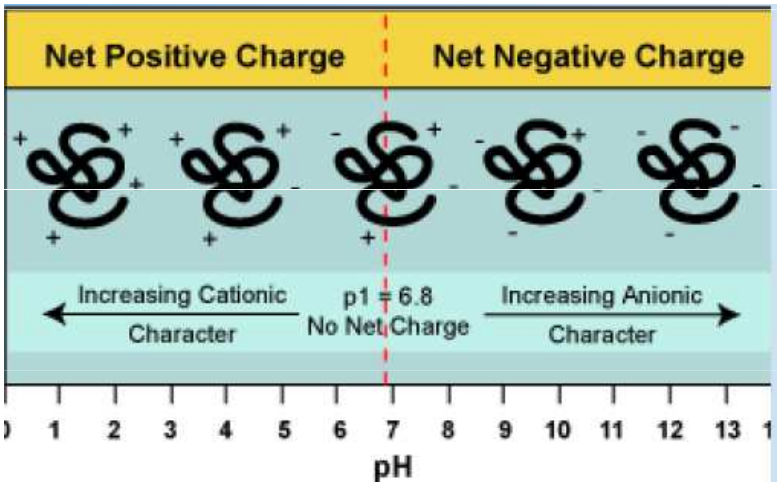

lysis of cells
centrifugation
chromatography
protein
Binding
Elution
metal ion affinity
Low
Molecular Weight
High
size exclusion
TEV cleavage
Elution of protein
Washing of contaminants
metal ion affinity
Binding
Elution
ion exchange

- Affinity chromatography - Immobilized metal ion affinity chromatography (IMAC)





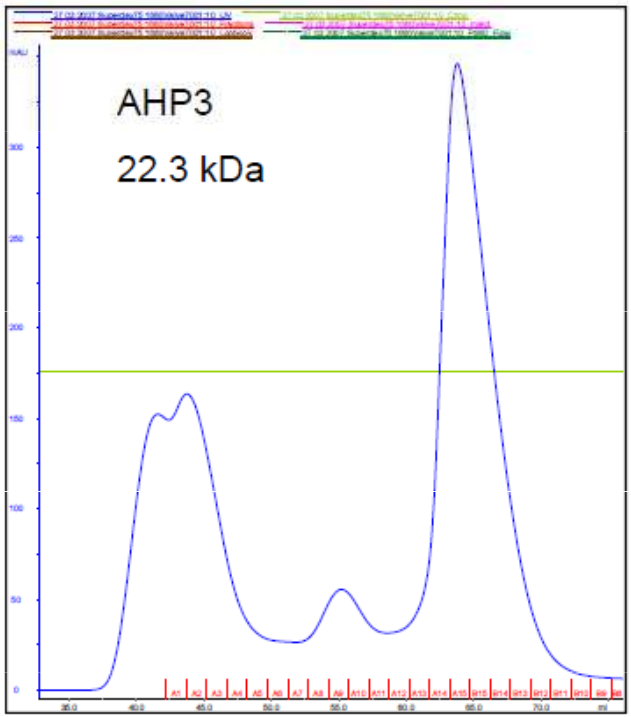
- Ion-exchange chromatography



- Optimal binding of recombinant protein with metal ion is achieved at pH 7–8.
- Buffers with a high salt concentration (0.5–1 M NaCl) reduce nonspecific electrostatic interaction.
- Elution of contaminating proteins can be achieved by lowering the pH or using low concentrations of imidazole.
- Elution of tagged protein is achieved at high imidazole concentrations (0–0.5 M), by strongly decreasing the pH, or by using EDTA.

Size-exclusion chromatography (Gel filtration)

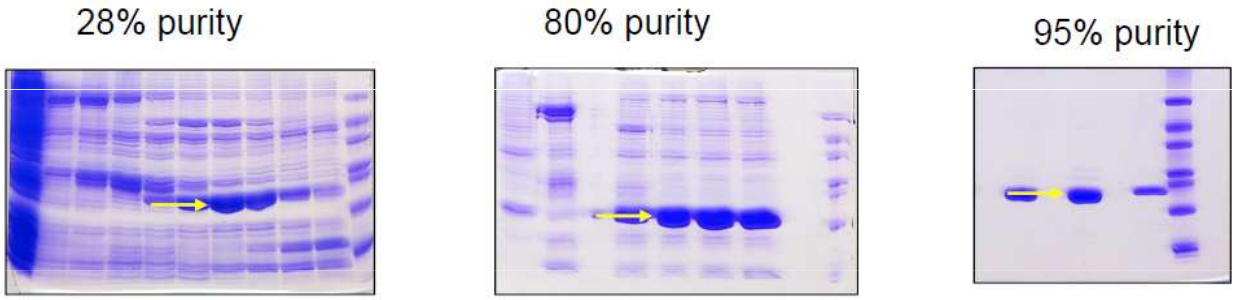
- porous beads
- Size-exclusion chromatography separates proteins
  on the basis of size



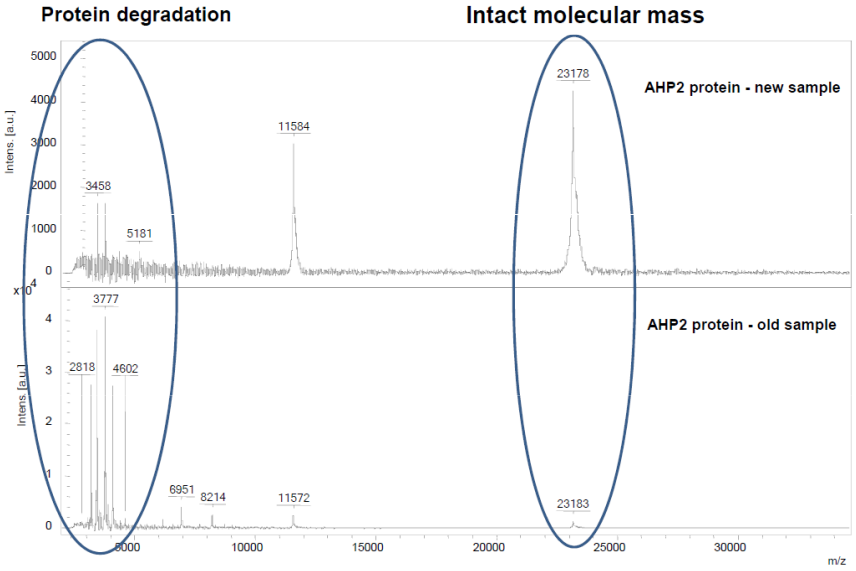
Protein Purity

What is molecular weight of 14-3-3zeta homodimer?

- SDS-PAGE



- Mass spectrometry (MS)-intact analysis, e.g. by MALDI-TOF MS



Another properties to check:

- secondary structure (e.g. CD)
- thermostability (typically in terms of Tr
- oligomeric state (e.g. DLS, AUC)

- Viac detailne a hlavne prakticke informacie na danu temu - v ramci predmetu: **C8980** Příprava a charakterizace proteinů I - Exprese a purifikace