Kód předmětu: Bi8980

MASARYKOVA UNIVERZITA

Protein expression and purification

- V. Protein expression

Lubomír Janda, Blanka Pekárová and Radka Dopitová

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

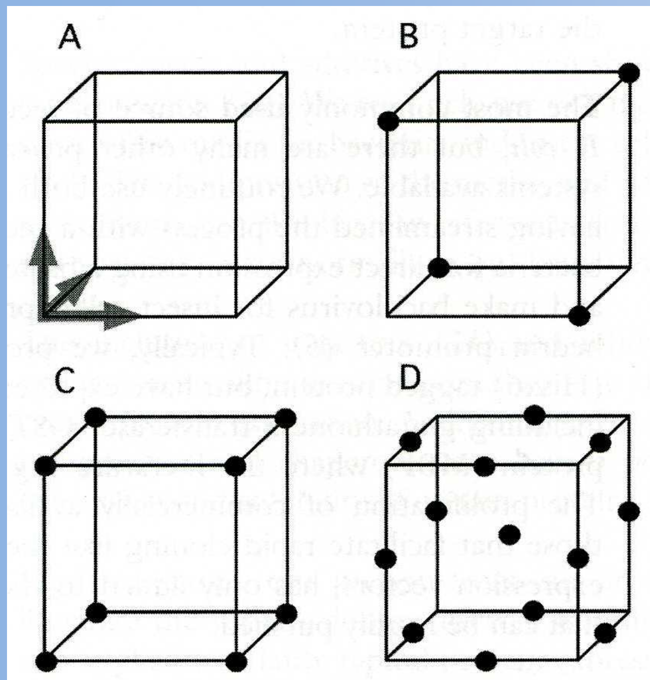
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

5.1. Designing experiments for high-throughput protein expression

High-throughput platform requires:

- Automation
- Miniaturization
- Quantitative management tools (to identify trends and relationships)

Experimental design:



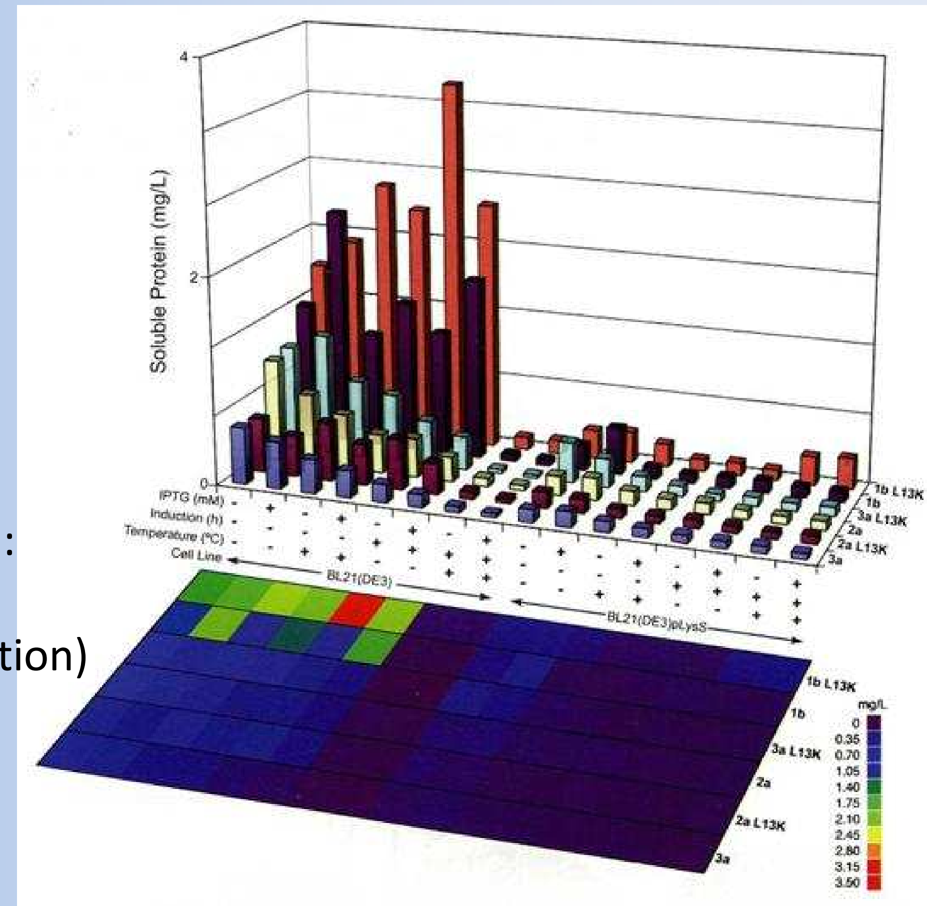
- An ill-defined experiment will often produce ambiguous results and fail to reach any conclusion.
- Analysis of quantitative response allows the experimenter to optimize conditions critical to production of a soluble protein.
- Performing one-factor-at-a-time experiments raises the risk of locating a local maximum (missing the actual best conditions).

A: One factor at a time
B: Fractional factorial
C: Full factorial
D: Response surface model (Box-Behnken design for three factors)

5.1. Designing experiments for high-throughput protein expression

Factors affecting expression:

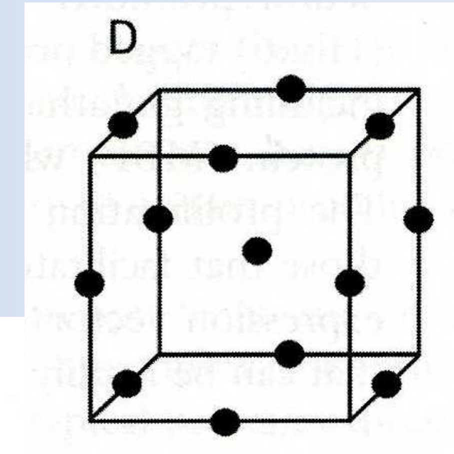
- Construct
- Expression system and vector
- Cell line (host strain)
- Temperature and time
- Media
- Additives



Full factorial design (16 conditions per construct):

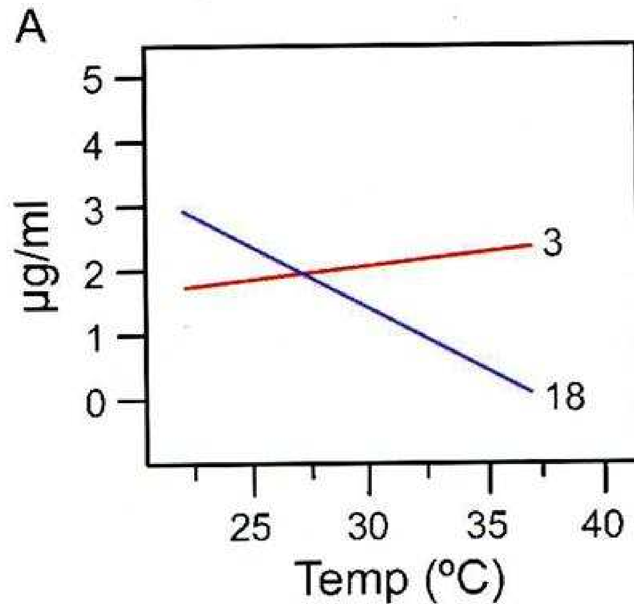
- three continuous factors (temperature, time, IPTG concentration)
- one categorical (host strain)

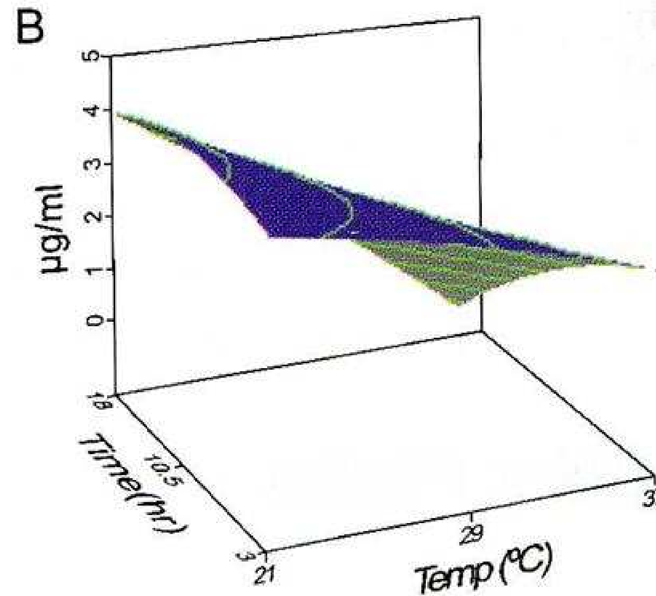
5.1. Designing experiments for high-throughput protein expression



Response surface model:

- fine-tunes the conditions
- capable to identify minimum or maximum





Design of experiment is merely a statistical tool, a means to an end.

It does not guarantee success and cannot replace technical expertise or creativity in experimental work.

5.2. Approaches for efficient protein production

I. Genetic approach x ***protein knowledgebase*** (biochemical approach)

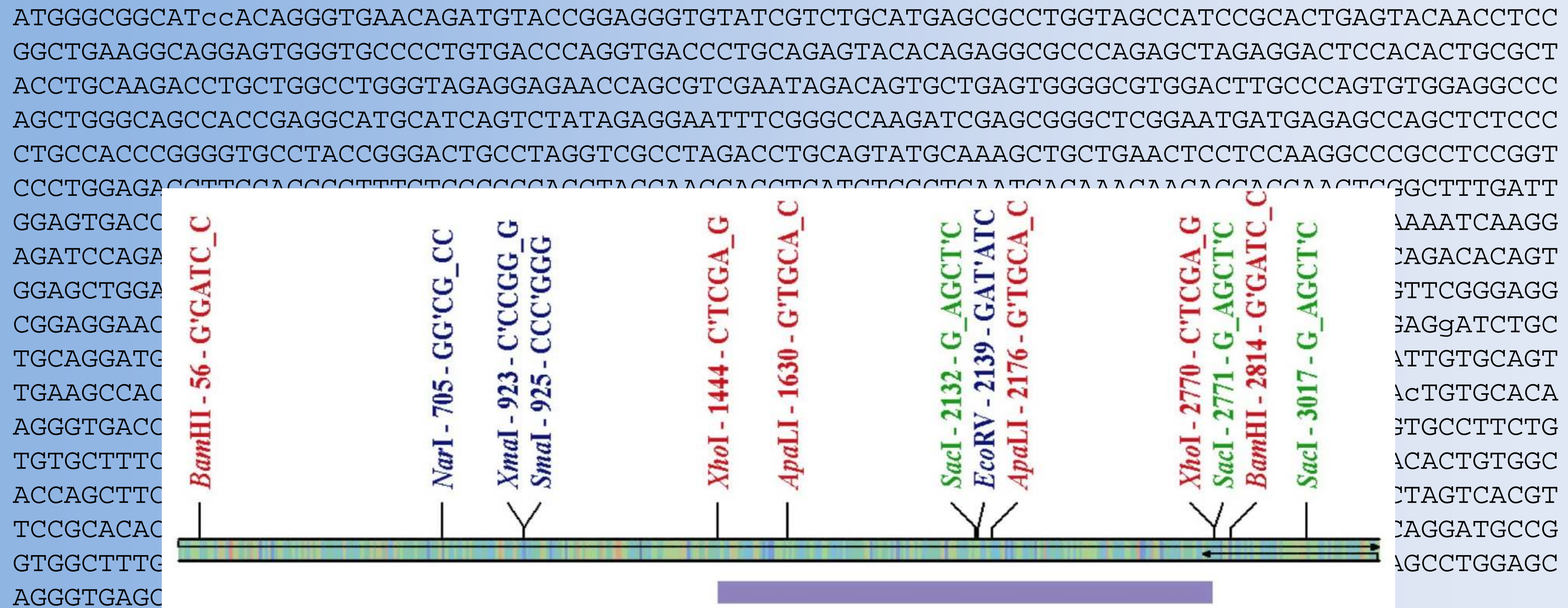
II. Expression density x ***functional activity***

III. Expression system x ***medium engineering***

IV. Troubles with removing tag fusion proteins x ***less convenient purification with classical chromatography***

5.2. Approaches for efficient protein production

5.2.1. Genetic approach x *protein knowledgebase*



DGVRANELQLRWQEYRELVLLLLQWIRHHTAAFEERKFPSSFEEIEILWCQFLKFKETELPAKEADKNRSKVIYQSLEGAVQAGQLK
IPPGYHPLDVEKEWGKLHVAILEREKQLRSEFERLECLQRIVSKLQMEAGLCEEQLNQADALLQSDIRLLASGKVAQRAGEVERDLD
KADGMIRLLFNDVQTLKDGRHPQGEQMYRRVYRLHERLVAIRTEYNLRLKAGVGAPVTQVTLQSTQRRPELEDSTLRYLQDLLAWVE
ENQRRIDSAEWGVDLPSVEAQLGSHRGMHQSIEEFRAKIERARNDESQLSPATRGAYRDCLGRLDLQYAKLLNSSKARLRSLESLHG
LQLCCCIEAHLKENTAYFQFFSDVREAEEQLQKLQETLRRKYSCDRTITVTRLEDLLQDAQDEKEQLNEYKGHLSGLAKRAKAIVQL
VEECQKFAKQYINAIKDYELQLITYKAQLEPVASPAKKPKVQSGSESVIQEYVDLRTRYSELTTLTSQYIKFISETLRRMEEEE

5.2. Approaches for efficient protein production

5.2.1. Genetic approach x ***protein knowledgebase***

Histidine kinase from *A. thaliana* - CKI 1

TM doména

HATPase

REC

Two constructs of receiver domain.

pET 28

pREC

REC

REC

Petra Borkovcová

Blanka Pekárová

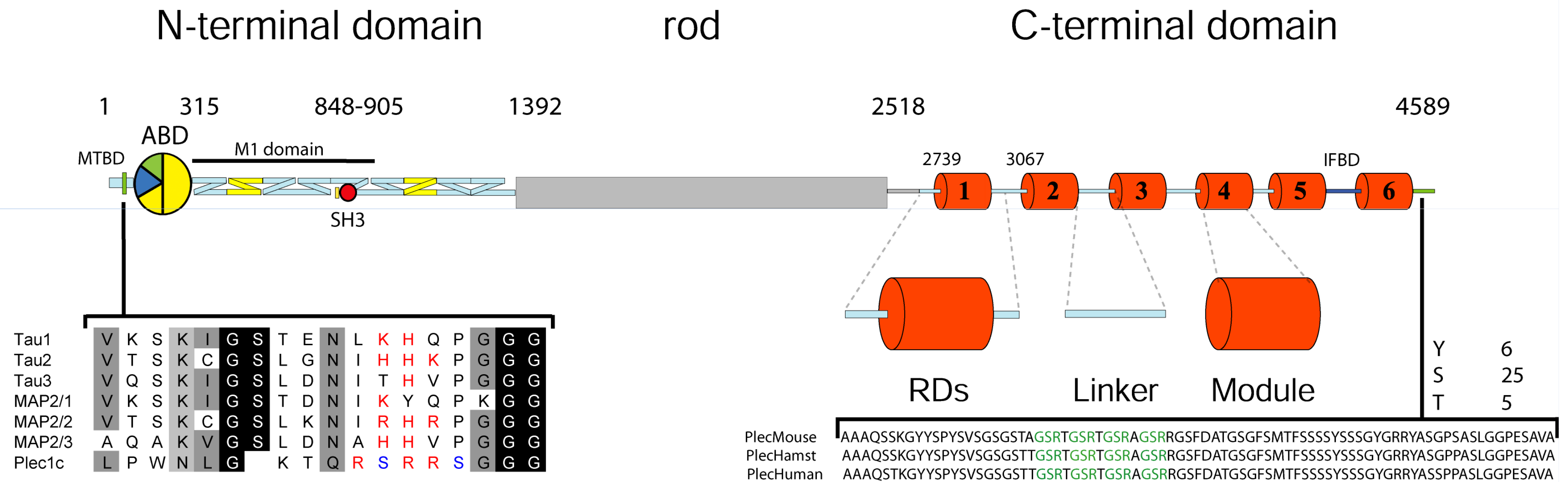
Lubica Urbaníková

Lubomír Janda

5.2. Approaches for efficient protein production

5.2.1. Genetic approach x ***protein knowledgebase***

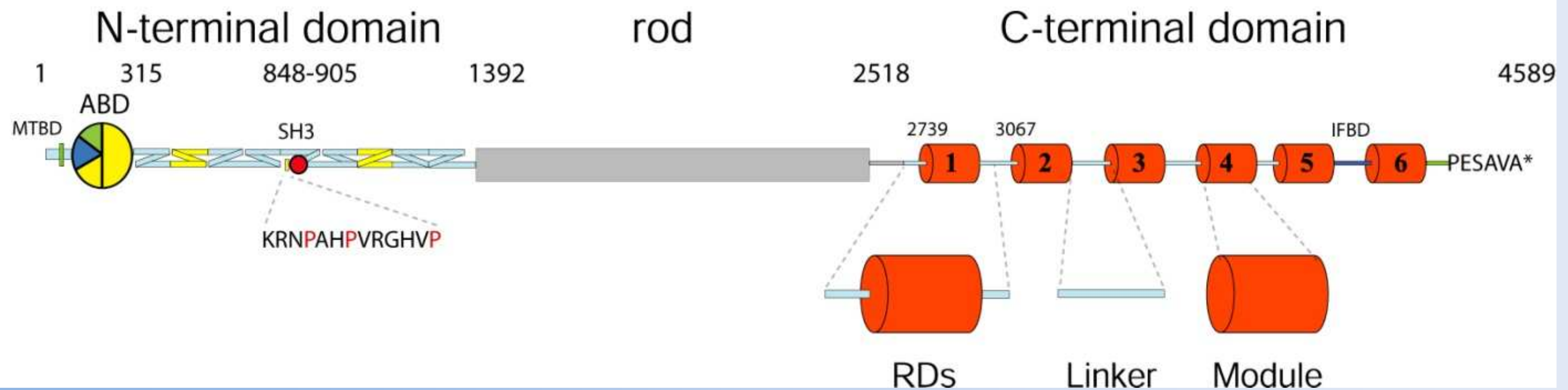
Plectin



Kd of Plectin (Ex 1-24) for actin	320 nM	(Ex 2-8)	25uM
Kd of Plectin (R5) for vimentin (IF)	100 nM		
Kd of Plectin (Ex 2-8) for integrin beta 4	170 nM		
Kd for microtubules in case of MAP2	1-3 uM		

5.2. Approaches for efficient protein production

5.2.1. Genetic approach x ***protein knowledgebase***



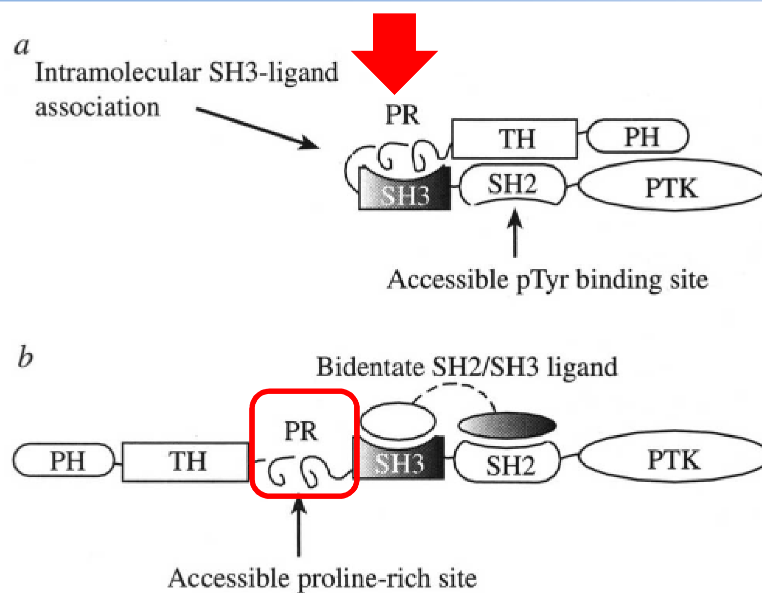
SH3 domain

Protein	Start	Sequence	End
Spectrin	978	VIALYDFEARSRREVSMKKNDVLTLLSSINKDWWKVEADDHQG FVPAVYVRKLA	1037
Actinin DM	834	VKSLFPFE GQG MKMDKGEVMLLKSKTNDDWWCVRKDNGVEGFVPANYVREVE	890
Itk Tyr kinase	167	VIALYDYQTNDPQELALRCDEEYYLLDSSEIHWWRVQDKNGHEGYAPSSYLVEKS	240
Envoplakin M	404	VDSICDWD SGEVQLLRGERCTLKDNADPYTWLVQGPGGETKSAPAACLWIPA	973
Actinin CE	847	VTALCDYS DENVTIKAGDDVYLLDNSDLIKWTIRDISGAEGQVP SVVFRI	914
Periplakin M	391	VEALCDFE GEQGLISRGYSYTLQKNN GESWELMDSAGNKLIAP AVCFVI	456
Kakapo DM	793	VQAICAYK QQGQLQIEKGETVTLLDNSGRVKWRVRTAKGQEGPIP GACLLL	862
Kakapo AG	941	VQSICSYK QGNISLDKNETCTLLDTSGRVKWRVKTSKGVEGSVH GVCLLL	1008
Desmoplakin M	536	LRALCDYK QDQKIVHKGDECILKDNNERSKWYVTGPGGVDMLVP SVGLII	604
MACF	858	VKAICDYR QIEITICKNDECVLEDNSQRTKWKVISPTGNEAMVP SVCFLI	926
Dystonin M	877	IKAICDYR QIEITIYKDDECVLANNSHRAKWKVISPTGNEAVVP SVCFTV	944
Plectin M	931	LIAVCDYK QVEVTVHKGDQCQLVGPAQPSHWKVLSGSSSEAAVP SVCFLV	1000

5.2. Approaches for efficient protein production

5.2.1. Genetic approach x ***protein knowledgebase***

```
                            PRD domain                 SH3 domain                                                   PRD
Spectrin        978                        AR  VIALYDFEARSRREVSMKKNDVLTLLSSINKDWWKVEADDHQG FVPAVYVRKLA  PDE L  1037
Actinin DM      834                        PH  VKSLFPFE  GQG MKMDKGEVMLLKSKTNDDWWCVRKDNGVEGFVPANYVREVE  PRP V   890
Itk Tyr kinase  167  KKPLPPT PEDNRRSFQEPEETL  VIALYDYQTNDPQELALRCDEEYYLLDSSEIHWWRVQDKNGHEGYAPSSYLVEKS  PNN L   240
Envoplakin M    404  PLPQRRN PSKQPLH           VDSICDWD   SGEVQLLRGERCTLKDNADPYTWLVQGPGGETKSAPAACLWIPA  PDP E   973
Actinin CE      847  PLWQRGERIPH PIK           VTALCDYS   DENVTIKAGDDVYLLDNSDLIKWTIRDISGAEGQVP  SWVFRI  PPT D   914
Periplakin M    391  LKYRRET PLK PIP           VEALCDFE   GEQGLISRGYSYTLQKNN GESWELMDSAGNKLIAP  AVCFVI  PPT D   456
Kakapo DM       793  PLNKRRQ PVNRQGP           VQAICAYK  QQGQLQIEKGETVTLLDNSGRVKWRVRTAKGQEGPIP  GACLLL  PPP D   862
Kakapo AG       941  PLKQRKQ PVNRQCT           VQSICSYK   QGNISLDKNETCTLLDTSGRVKWRVKTSKGVEGSVH  GVCLLL  PPP D  1008
Desmoplakin M   536  LKPRNPDYRSNKPII           LRALCDYK   QDQKIVHKGDECILKDNNERSKWYVTGPGGVDMLVP  SVGLII  PPP N   604
MACF            858  LKPRNPDHVLKSTLS           VKAICDYR   QIEITICKNDECVLEDNSQRTKWKVISPTGNEAMVP  SVCFLI  PPP N   926
Dystonin M      877  LKPRNPDNPLKTSIP           IKAICDYR   QIEITIYKDDECVLANNSHRAKWKWISPTGNEAVVP  SWCFTV  PPP N   944
Plectin M       931  LKPRNPAHPVRGHVP           LIAVCDYK   QVEVTVHKGDQCQLVGPAQPSHWKVLSGSSSEAAVP  SVCFLV  PPP N  1000
```



a) Model of observed intramolecular interaction showing the observed interaction between the Itk proline-rich region and SH3 domain.

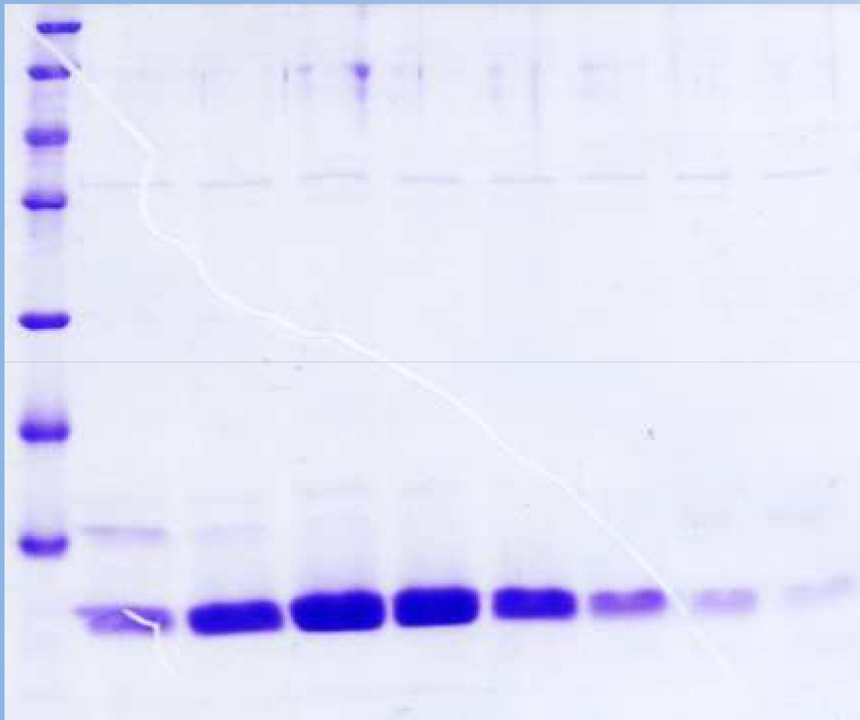
b) Model of the opening of the intramolecular complex by interaction with bidentate ligand for the Itk SH3 and SH2 domains.
(Andreotti et al., Nature 1996)

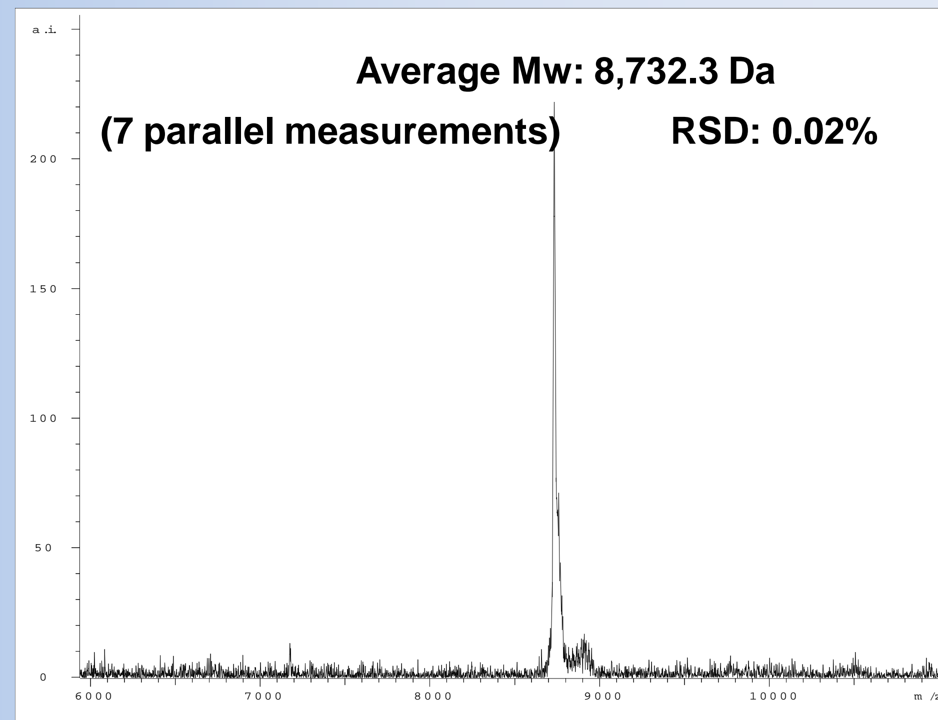
5.2. Approaches for efficient protein production

5.2.1. Genetic approach x ***protein knowledgebase***

SH3 domain of plectin with surrounding proline rich regions

(Sarc homology domain soluble in citrate buffer of pH 3.5)





Theoretical pI/Mw (average) for the protein sequence

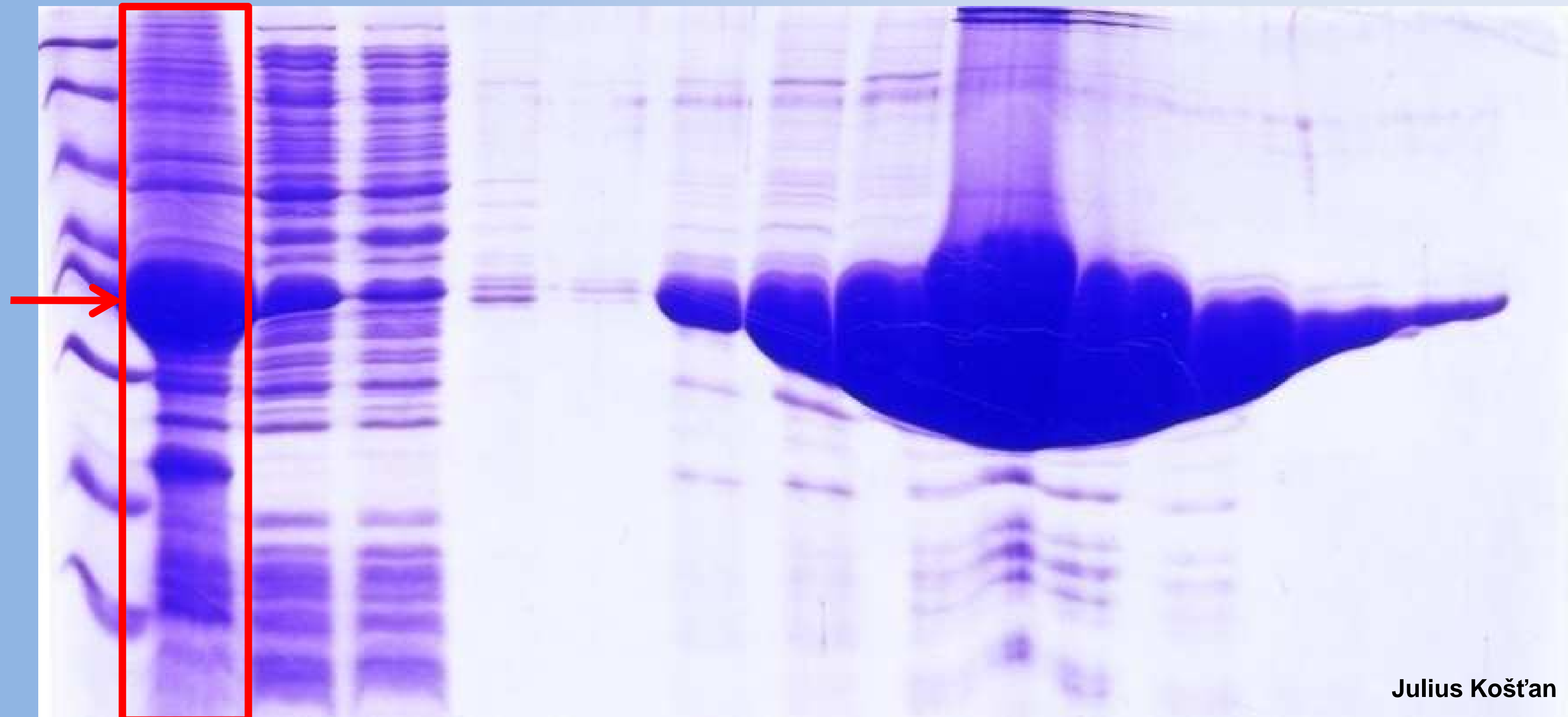
Theoretical pI/Mw: 7.78 / **8,726.11**

mefKAIVQLKPRNPAHPVRGHVPLIAVCDYKQVEVTVHKGD
QCQLVGPAQPSHWKVLSGSSSEAAVPSVCFLVPPPNQEf

5.2. Approaches for efficient protein production

5.2.2. Expression density x ***functional activity***

Expression and purification of plectin's ABD (Actin Binding Domain) in three isoforms.



5.2. Approaches for efficient protein production

5.2.2. Expression density x ***functional activity***

Maize recombinant β–glucosidase produced in *E. coli.*

Cultivation condition	***Yield (mg)***	***Specific activity (nkat/mg) /(total activity nkat)***
LB medium	*380*	*1.9* (966 nkat)
TB medium - pH 6	*230*	*3.8* (874 nkat)
TB medium – pH 7	*230*	*4.2!* (966) nkat)
TB medium – pH 8	*410*	*2.8* (1,148 nkat)
Additive of cellobiose (LB medium)	*400*	*2.7!* (1,080 nkat)

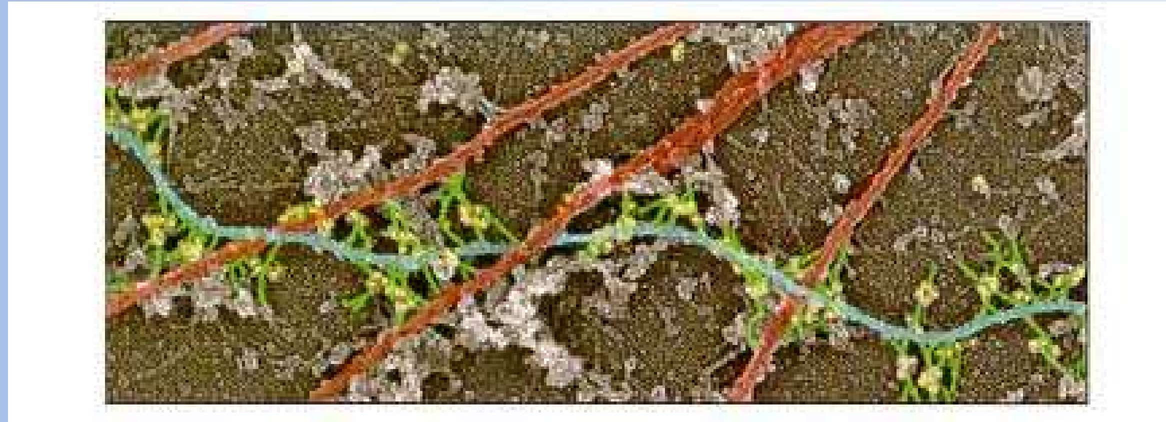
Radka Fohlerov

Result: TB medium (pH 7.0) supplemented by cellobiose shows 3.1 x higher β-glucosidase specific activity than in common LB medium.

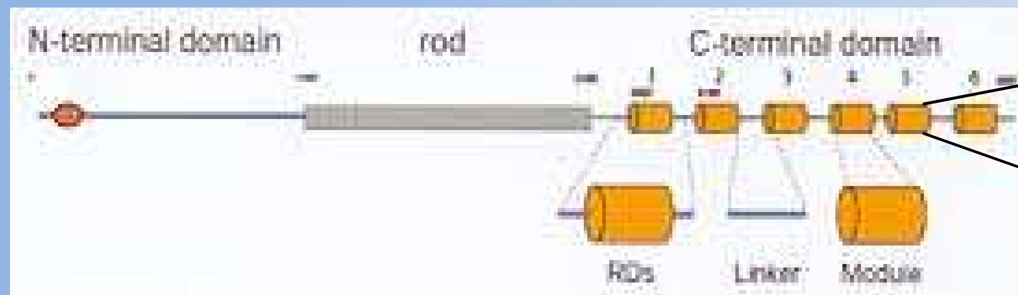
5.2. Approaches for efficient protein production

5.2.2. Expression density x ***functional activity***

The cytolinker protein:
plectin



Plectin is one of the main linker proteins for the cytoskeleton.



Plectin R5

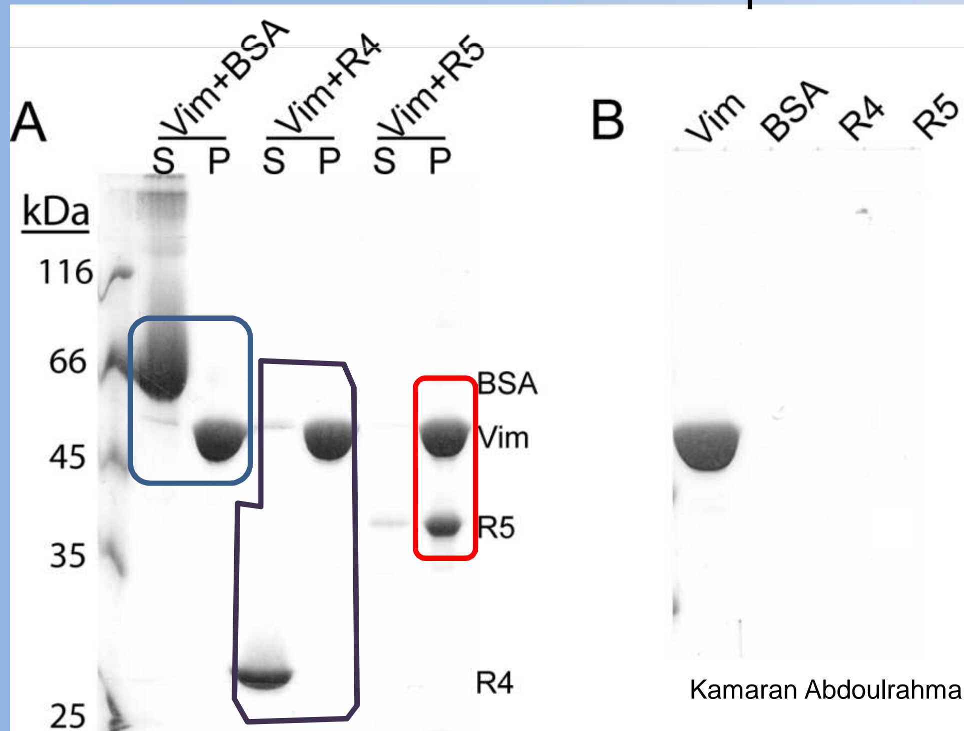
5.2. Approaches for efficient protein production

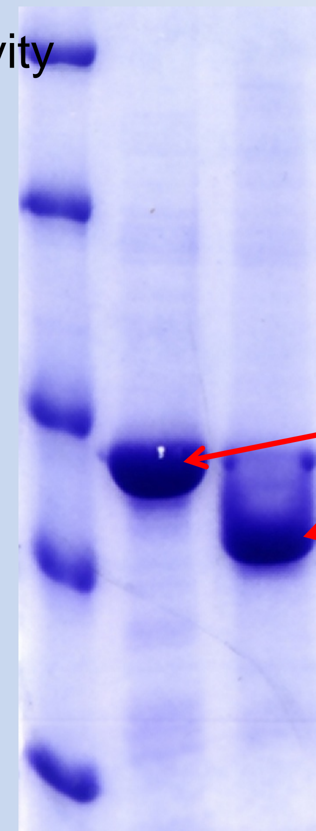
5.2.2. Expression density x ***functional activity***

Converted pET 15b + IF binding domain of plectin

• R5 d. plectin (pH 7.9)	Insoluble form
• R5 d. plectin (pH 7.9, urea, dialysis)	Func. act.(45%)
• R5 d. plectin (pH 7.9, urea, refolding HR)	Func. act.(60%)
• R5 d. plectin (pH 11, purification pH 9.0)	Func. act. (≥95%)

Co-sedimentation – functional test on protein activity



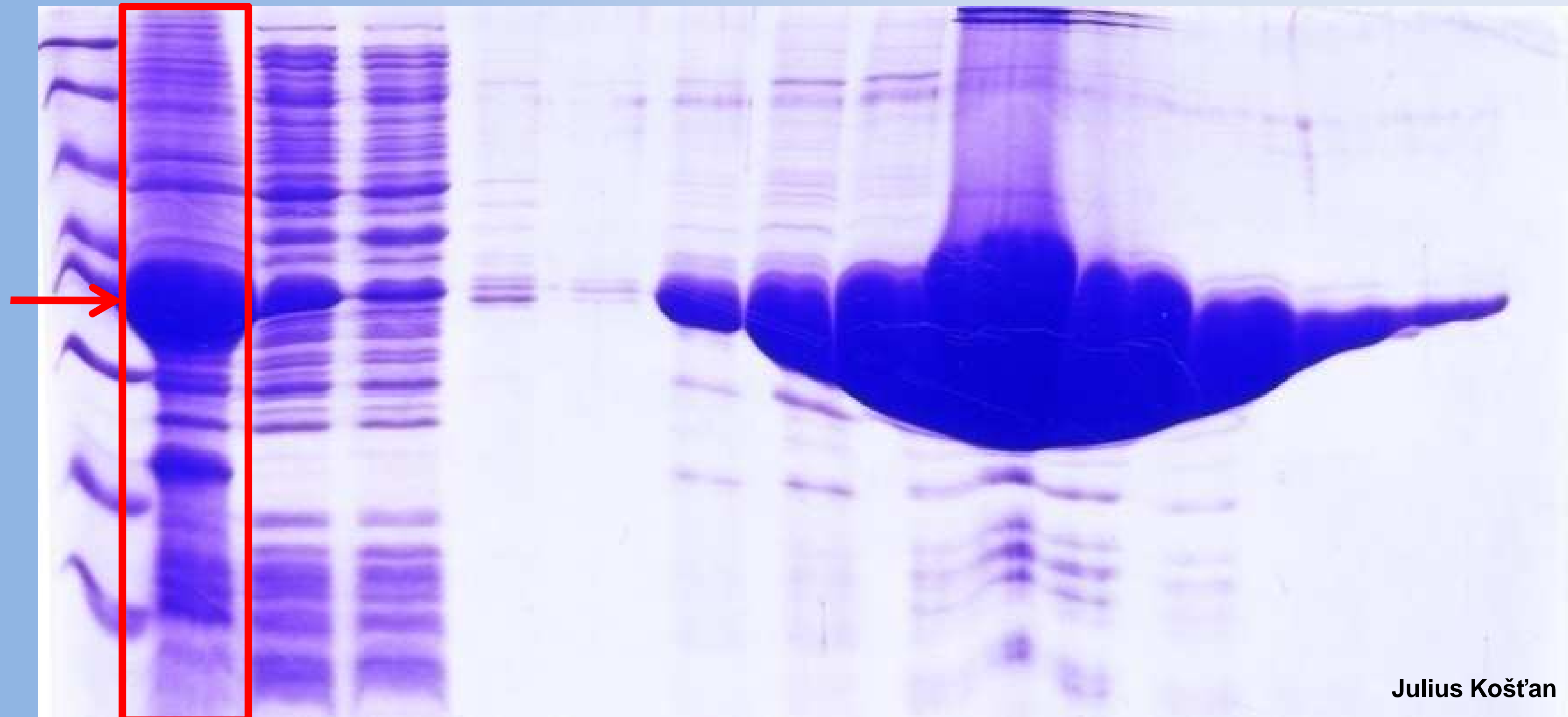


1. R5wt oxidased form + DTT
2. R5wt oxidased form – DTT

5.2. Approaches for efficient protein production

5.2.2. Expression density x ***functional activity***

Expression and purification of plectin's ABD (actin binding domain) in three isoforms.

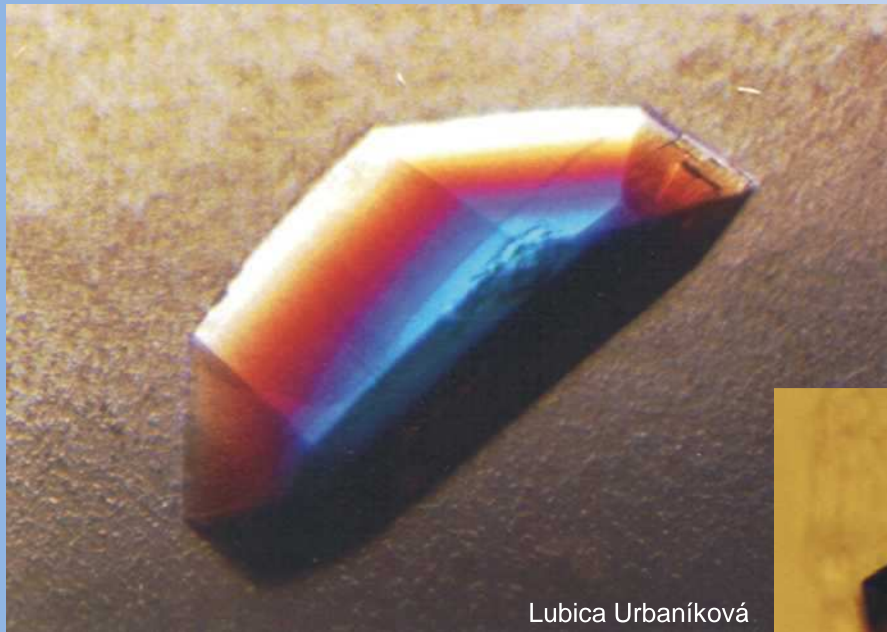


5.2. Approaches for efficient protein production

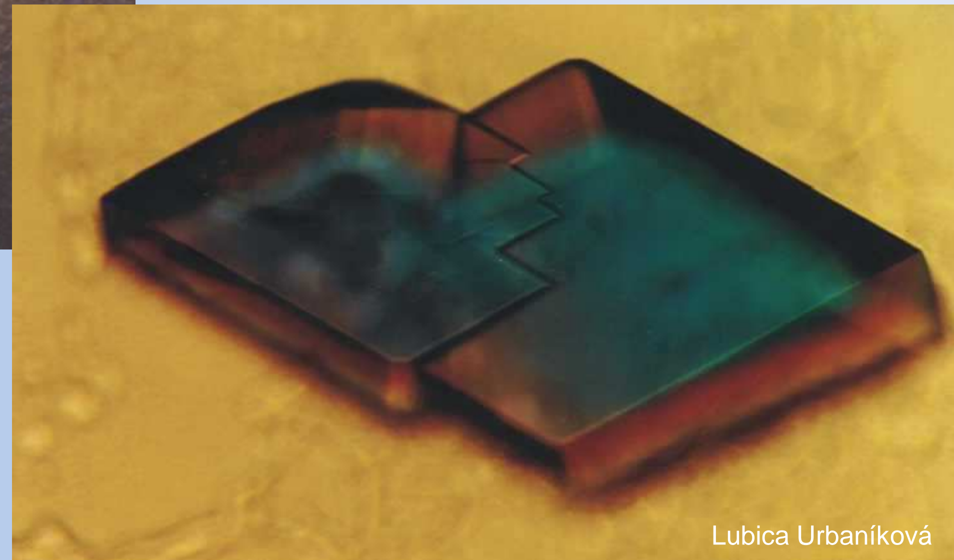
5.2.2. Expression density x ***functional activity***

Monoclinic crystals of plectin ABD

Precipitant solution: 0.1 M TRIS buffer pH 8.5
10% PEG 4000


Lubica Urbaníková

2% dioxane


Lubica Urbaníková

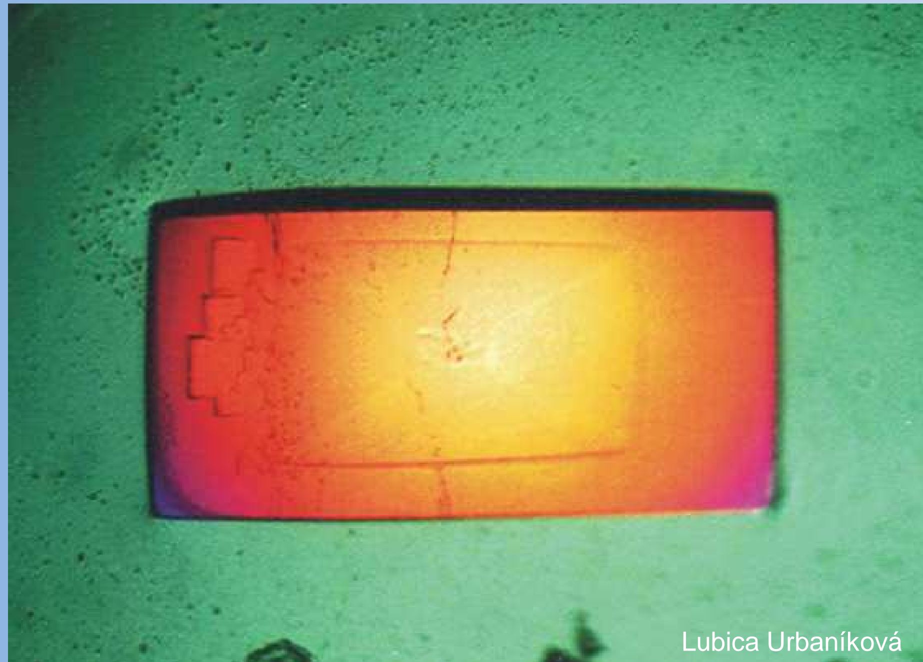
Space group $P2_1$
2 molecules in asym. unit

5.2. Approaches for efficient protein production

5.2.2. Expression density x ***functional activity***

Orthorhombic crystals of plectin ABD

Precipitant solution:



0.1 M Cacodylate buffer pH 6.5
6–8% PEG 8000
0.2 M Ca acetate
2% dioxane

Space group $P2_12_12_1$

1 molecule in asym. unit

2.0 Å resolution

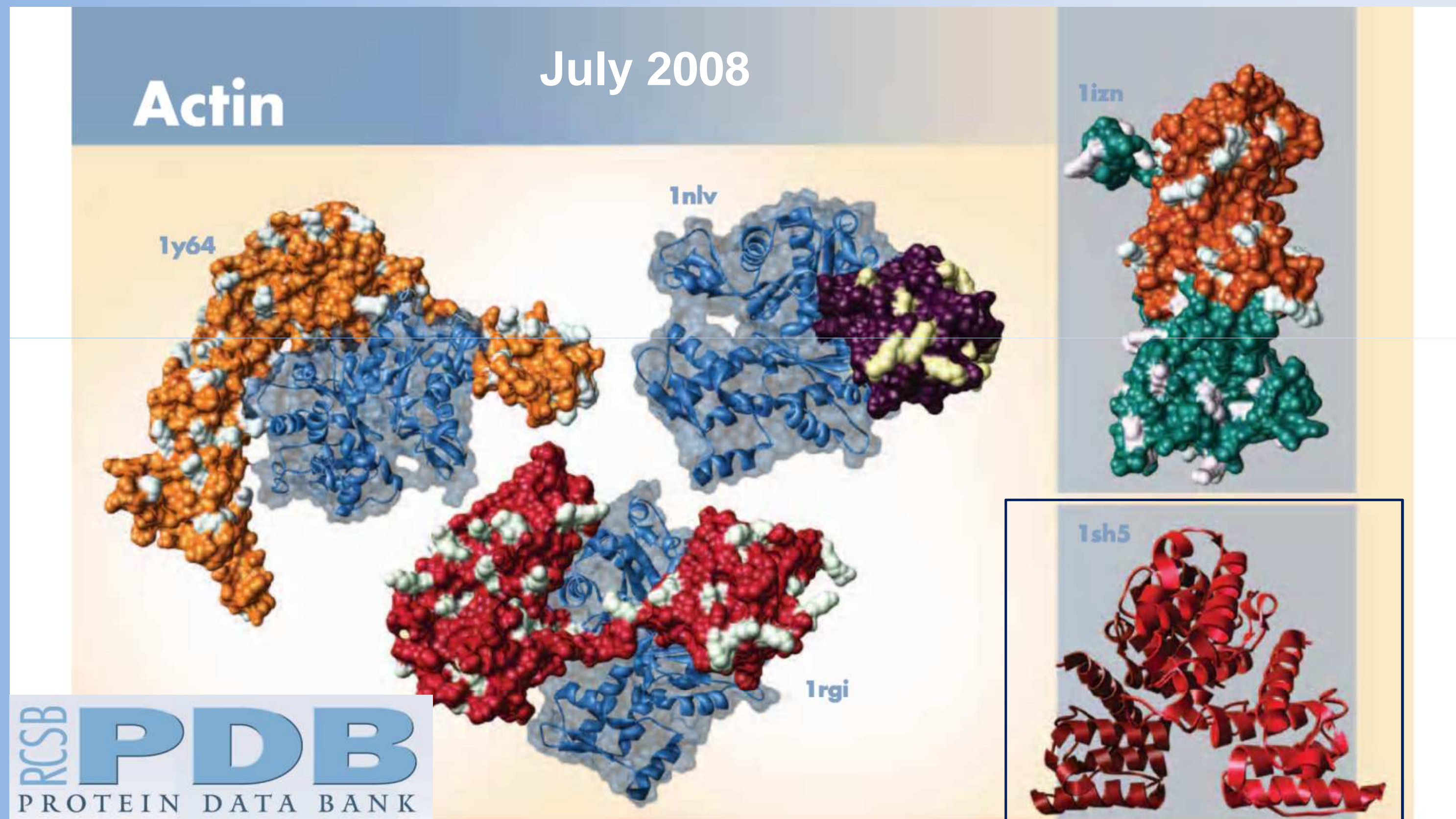
V. Protein expression

1nlv: S.M.Vorobiev, B. Strokopytov, D.G. Drubin, C. Frieden, S. Ono, J. Condeelis, P.A. Rubenstein, S.C. Almo. The structure of non-vertebrate actin: Implications for the ATP hydrolytic mechanism (2003). *Proc.Natl.Acad.Sci. USA* ***100:5760-5765.***

1rgi: L.D.Burtnick, D. Urosev, E. Irobi, K. Narayan, R.C. Robinson (2004). Structure of the N-terminal half of gelsolin bound to actin: roles in severing, apoptosis and FAF. *EMBO J.* **23:2713-2722.**

1izn: A.Yamashita, K. Maeda, Y. Maeda (2003). Crystal structure of CapZ: structural basis for actin filament barbed end capping. *EMBO J.* **22:1529-1538.**

1sh5: J. Sevcik, L. Urbanikova, J. Kostan, L. Janda, G. Wiche (2004). Actin-binding domain of mouse plectin: crystal structure and binding to vimentin. *Eur.J.Biochem.* ***271:873-*** **1884**.



5.2. Approaches for efficient protein production

5.2.3. Expression system x ***medium engineering***

Examples of *E. coli* expression systems and web pages for further information

	Vector system	Promoter/ induction method	Special host strains required	Protein tag	Source (website)
	Pinpoint	*tac*/IPTG or *T7* IPTG	Yes	Biotin binding domain	www.promega.com
*	pET	*T7* IPTG	Yes	His_6, T7 gene	http://www.merckbiosciences.co.uk
*	pGEX	*tac*/IPTG	No	GST	www.amershambiosciences.com
	pBAD	*ara*BAD	Yes	His_6, GFP	www.invitrogen.com
	pLEX	P_L/trp	Yes		
*	pRSET	*T7* IPTG	Yes	His_6, T7	
	pPROTet	P_{Ltet}/anhydrotetracycline	No	His_6	www.clontech.com
	pTYB	*T7* IPTG	Yes	Chitin binding domain	www.neb.com
*	pMAL	*tac*/IPTG	Yes	Maltose binding domain	
	pQE	*T5*/IPTG	Yes/TOPP	His_6	www.qiagen.com
	pCAL	*T7*/IPTG	Yes	Calmodulin binding peptide	www.stratagene.com
	pFLAG	*tac*/IPTG	Yes		www.sigmaaldrich.com

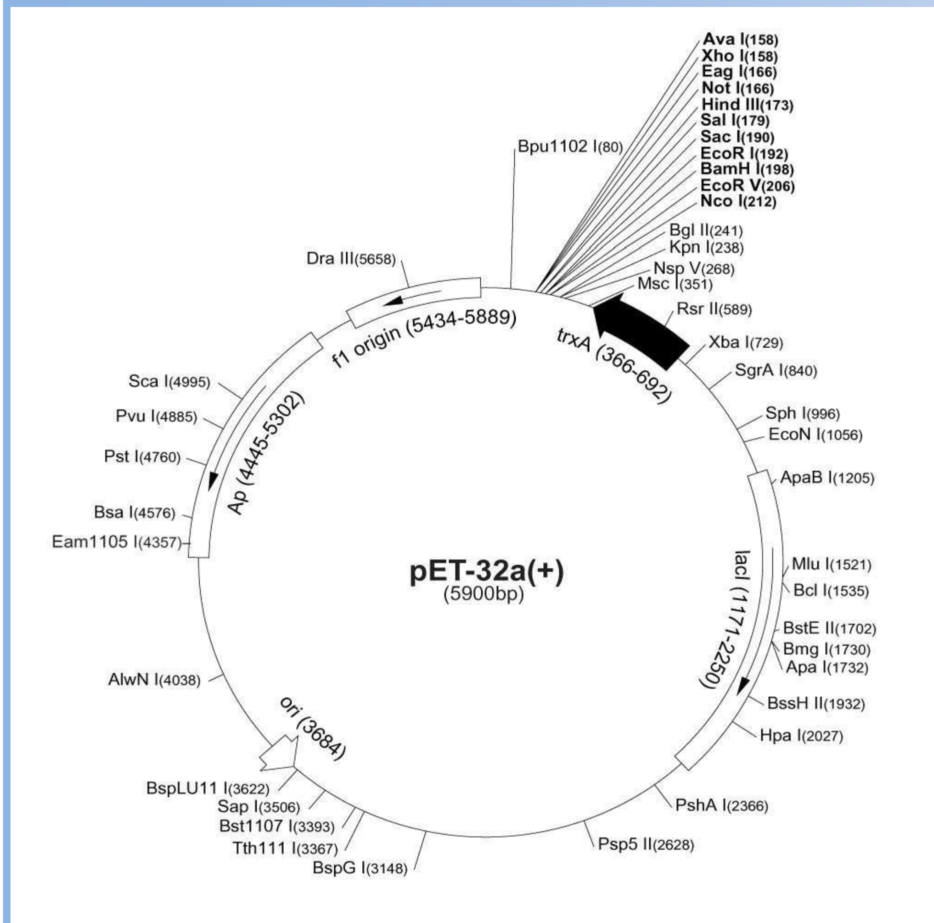
5.2. Approaches for efficient protein production

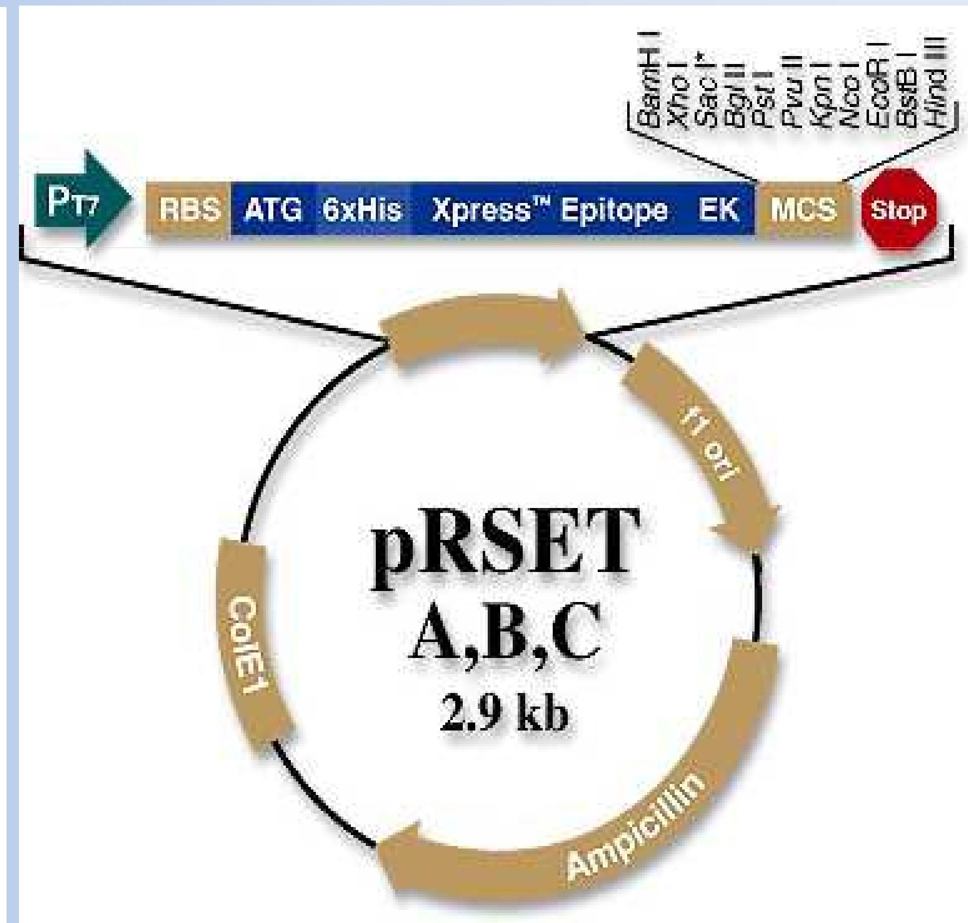
5.2.3. Expression system x ***medium engineering***

pET32a::AHP1
pET32a::AHP5

AHP – phosphotransfer protein in cytokinin signalling pathway of *A. thaliana*

pRSETB::AHP1
pRSETB::AHP5





5.2. Approaches for efficient protein production

5.2.3. Expression system x ***medium engineering***

5.2.3.1. Temperature

Expression system

AHP1

Plasmid	**pET32a+**		**pRSETB**	
Temperature (℃) growth/induction	Soluble form (%)	Insoluble form (%)	Soluble form (%)	Insoluble form (%)
22℃/22℃	62%	38%	71%	29%
37℃/22℃	0%	100%	82%	18%
37℃/28℃	0%	100%	8%	92%

AHP5

Plasmid	**pET32a+**		**pRSETB**	
Temperature (℃) growth/induction	Soluble form (%)	Insoluble form (%)	Soluble form (%)	Insoluble form (%)
22℃/22℃	78%	22%	76%	24%
37℃/22℃	67%	33%	81%	19%
37℃/28℃	61%	39%	81%	19%

Radka Fohlerová

Lubomír Janda

5.2. Approaches for efficient protein production

5.2.3. Expression system x ***medium engineering***

5.2.3.1. Temperature

Production of soluble AHP proteins using *E. coli* expression vector pRSET at different temperatures (%)

Temperature (°C) growth/induction	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%

Radka Fohlerová

5.2. Approaches for efficient protein production

5.2.3. Expression system x ***medium engineering***

5.2.3.2. Medium pH

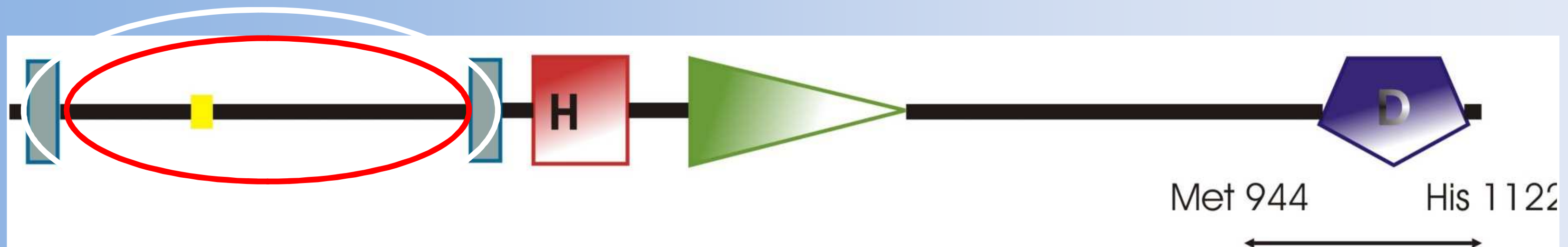
Production receiver domain of plant histidine kinase AHK4 in *E. coli* by *pET161DEST*			
pH	6.0	7.0	8.0
Soluble fraction	35%	89%	100%

Radka Fohlerová

- **Disintegrate** *E. coli* in native buffer.
- **Denaturated** crude extract by chaotropic compounds (urea).
- Load **SDS-PAGE.**
- **Scan** the gel after staining and subsequent de-staining.
- Evaluate differences between signals from protein denaturated by chaotropic compounds and protein signal from native buffer.

5.2. Approaches for efficient protein production

5.2.3. Expression system x ***medium engineering***



pDEST17::CKI1ex1 – 371 AA, Mw = 42 kDa
pDEST17::CKI1ex2 – 419 AA, Mw = 47 kDa

5.2. Approaches for efficient protein production

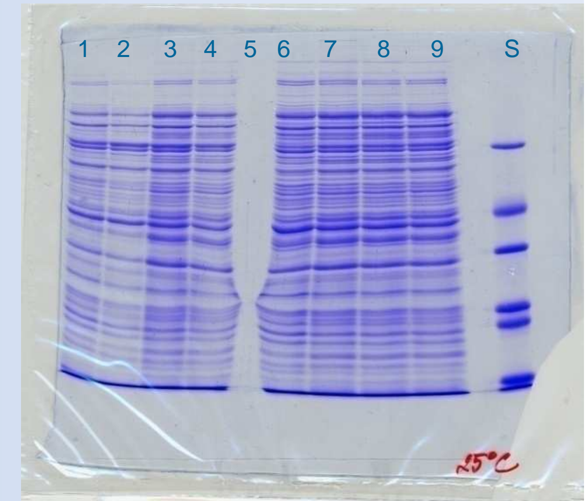
5.2.3. Expression system x ***medium engineering***

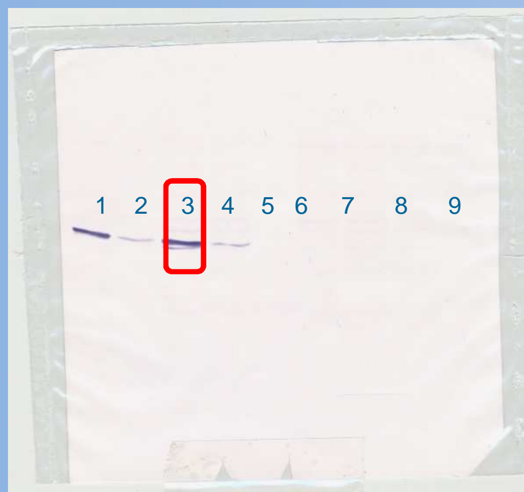
Growth temperature
37℃, expression 28℃

Growth and expression 25°C

1: BL21 – 1 h after induction = OD 0.5
2: 4 h after induction = OD 0.5
3: 1 h after induction = OD 2
4: 4 h after induction = OD 2
5: -
6: C43 – 1 h after induction = OD 0.5
7: 4 h after induction = OD 0.5
8: 1 h after induction = OD 2
9: 4 h after induction = OD 2

S: 14–66 kDa





1: BL 21 before induction
2: 1 h after induction = OD 0.5
3: 3 h after induction = OD 0.5
4: 2 h after induction = OD 2
5: -
6: C43 before induction
7: 1 h after induction = OD 0.5
8: 3 h after induction = OD 0.5
9: 2 h after induction = OD 2

S: 14–66 kDa

Petra Borkovcová

5.2. Approaches for efficient protein production

5.2.3. Expression system x ***medium engineering***

Expression

TB medium: pH 6.0, pH 7.0,

growth → induction
growth 25℃ → induction 22℃, 13.5 h

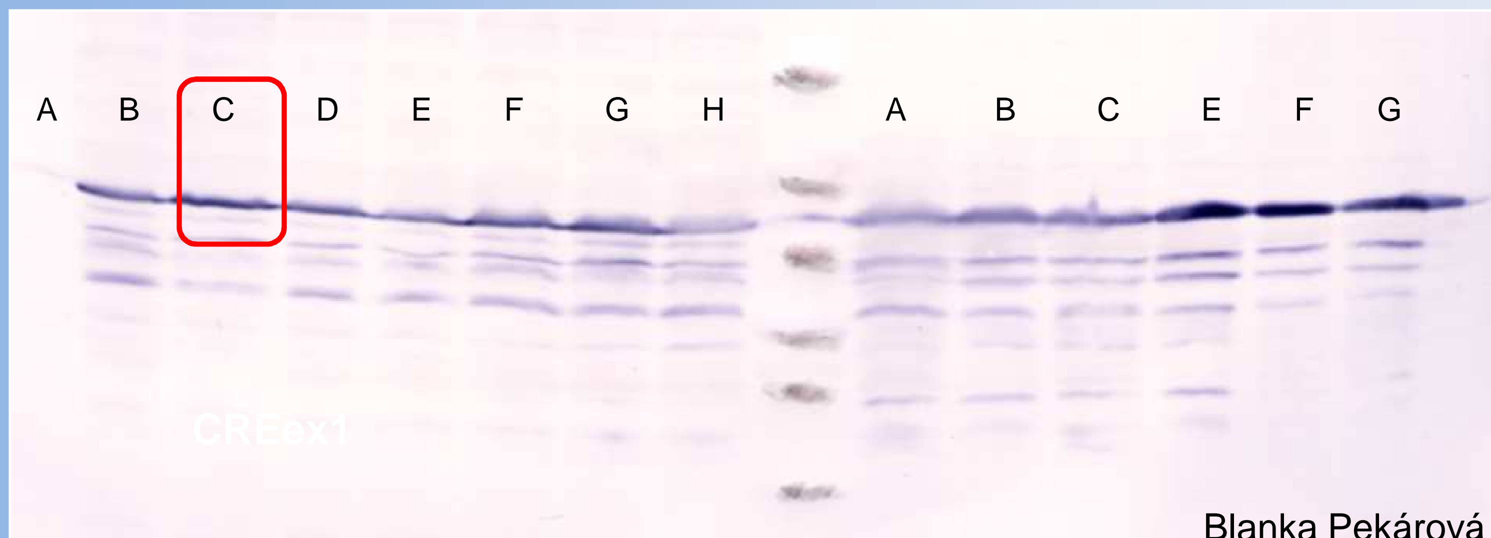
Lysis buffers:

A. Citrate b., pH 3.6, Triton X-100
B. Tris b., pH 7.9, Triton X-100
C.
D. Tris b., pH 7.9, Triton X-100
E. Tris b., pH 7.9, CTAB
F. Tris b., pH 7.9, NONIDET P-40
G. Tris b., pH 7.9, SDS
H. Tris b., pH 7.9

CKI1ex1

supernatant

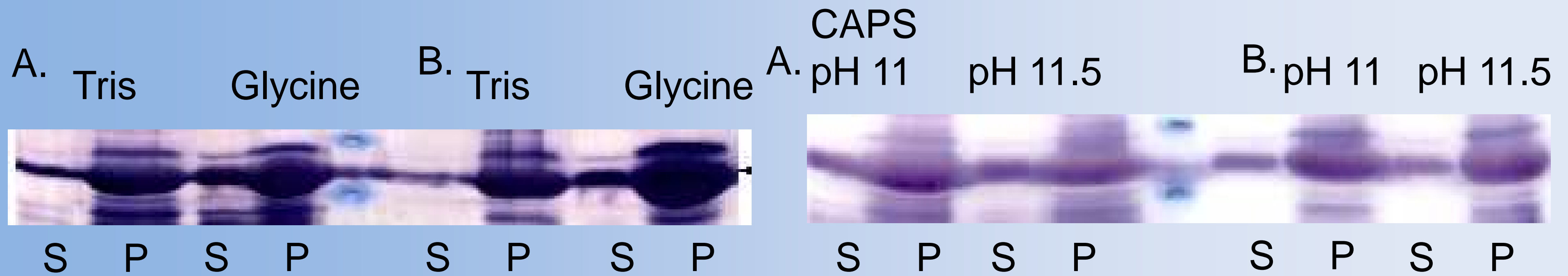
pellet



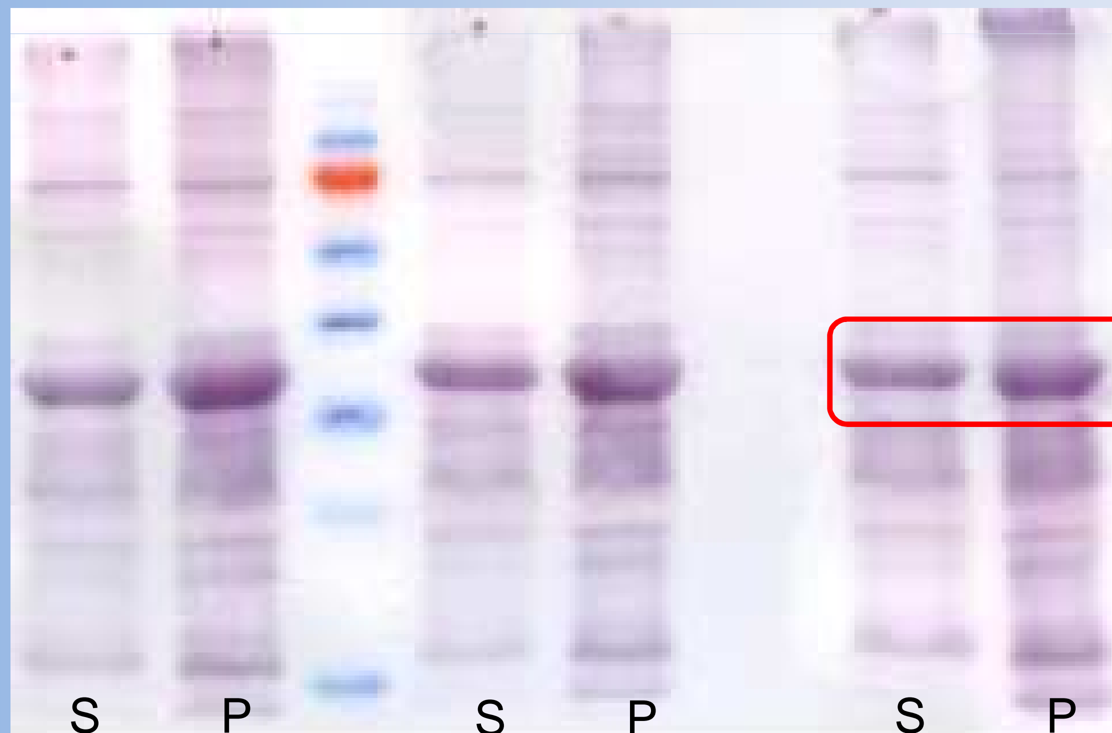
5.2. Approaches for efficient protein production

5.2.3. Expression system x ***medium engineering***

A. E. coli BL21(DE3) Arctica
B. E. coli BL21 (DE3)$_{RIL}$



Western blot detected by poly-His antibodies



Severine Jansen

5.2. Approaches for efficient protein production

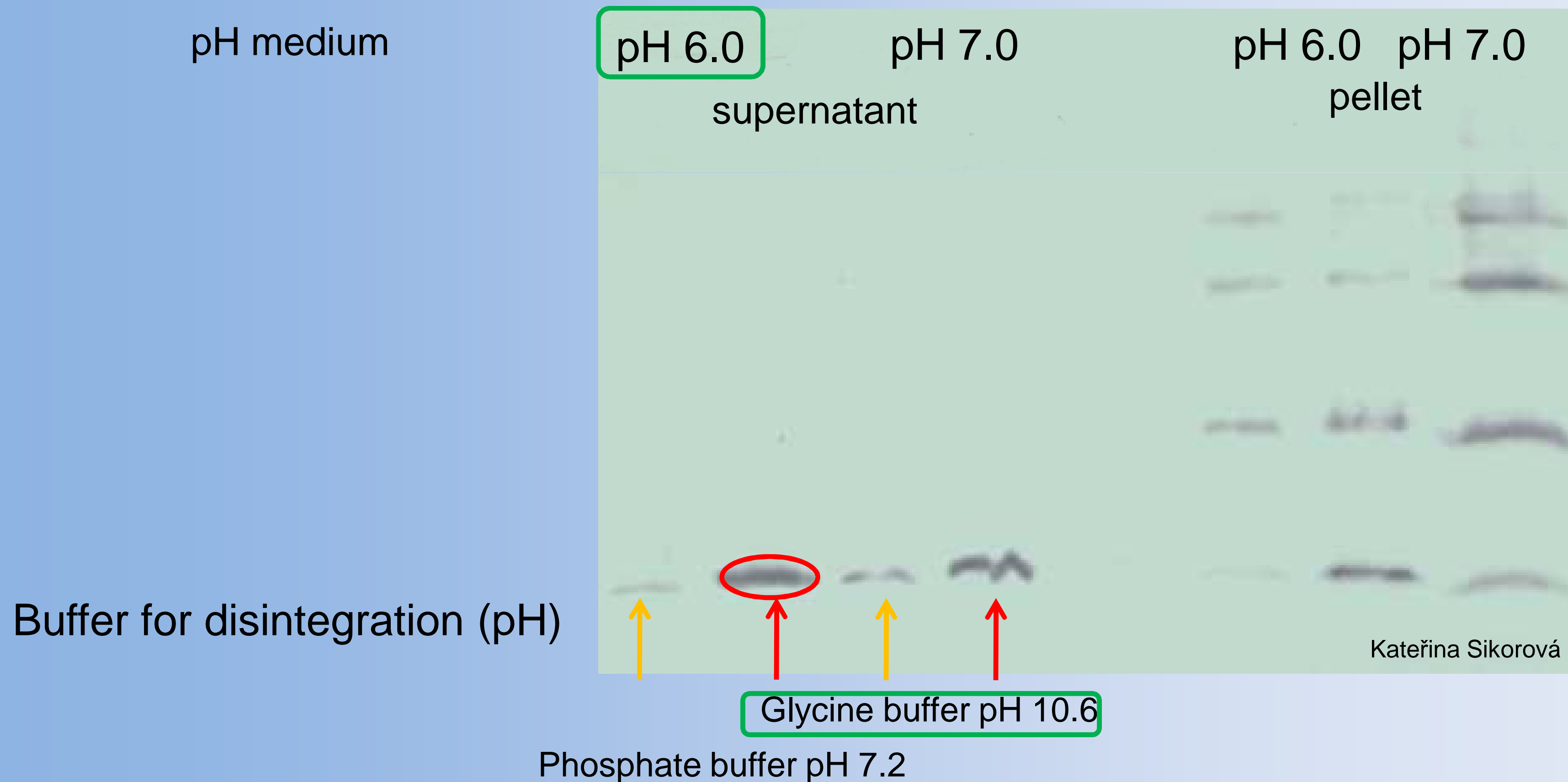
5.2.3. Expression system x ***medium engineering***

5.2.3.3. Buffer for disintegration (pH)

LTP-2 (non-specific lipid transporting protein from wheat)

pH medium

Buffer for disintegration (pH)



5.2. Approaches for efficient protein production

5.2.4. Troubles with removing tag fusion protein x less convenient purification with column

N-terminal amino acids which reduce protein stability

- Arg, Lys, Phe, Leu, Trp and Tyr
  - Tobias et al., 1991, Science

Amino acids in penultimate position (second behind N-terminal methionine) enhancing protein stability.

- His, Gln, Glu, Phe, Met, Lys, Tyr, Trp, Arg

  - Hirel et. al., 1989, PNAS and Lathrop et al. 1992

  - Liao et al., 2004, Protein Science

5.2. Approaches for efficient protein production

5.2.4. Troubles with removing tag fusion protein x less convenient purification with column

Control of protease cleavage sites in fusion proteins

- Thrombin — Pro-Arg/Gly, Pro-Lys/Leu, Ala-Arg/Gly, Gly-Lys/Ala, Ile-Arg/Ser, Leu-Arg/Ala, Ile-Arg/Ile
  pH 8.0
- PreScission — Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro
  pH 8.9
- Factor Xa — Ile-Glu-Gly-Arg/X
  pH 6.5–7.5
- Enterokinase — Asp-Asp-Asp-Asp-Lys/X
  pH 7.0–8.0

~~AHP5~~

5.2. Approaches for efficient protein production

5.2.5. Maximizing target protein recovery

- Protein expression
- Protein purification

- Take advantage of protein knowledgebase for construction of expression vector
- Medium engineering
  - protein induction at different temperature
  - medium pH
  - additives
  - buffer options for disintegration
  - degas all the buffers
- Test on protein activity
- Encourage protein purification without tag sequences

5.3. Expression system

5.3.1. *E. coli* expression system

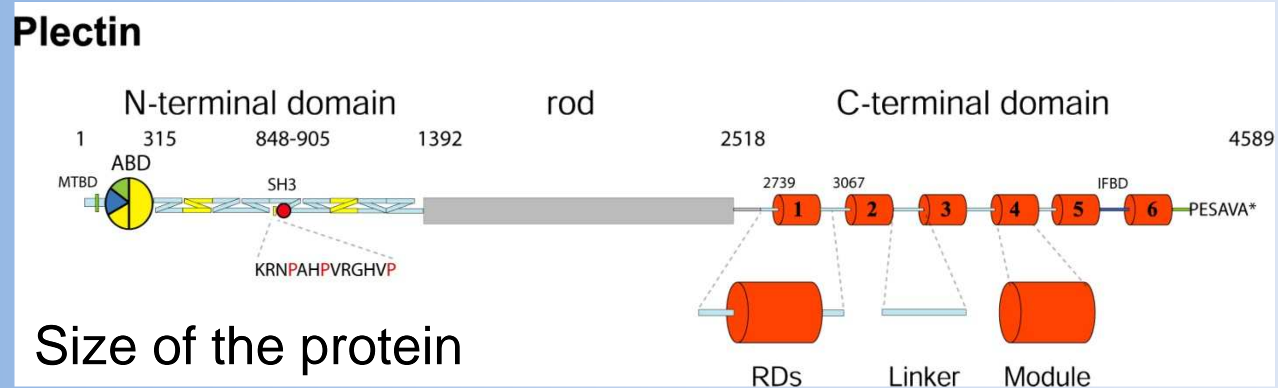
Advantages and Disadvantages of *E. coli*

- Ease of gene manipulation
- Availability of reagents
- Easy of producing quantities of protein
- Speed
- Low cost
- Adaptability of the system

- Formation of insoluble inclusion bodies
- Size of the protein
- Post-translational modification

5.3. Expression system

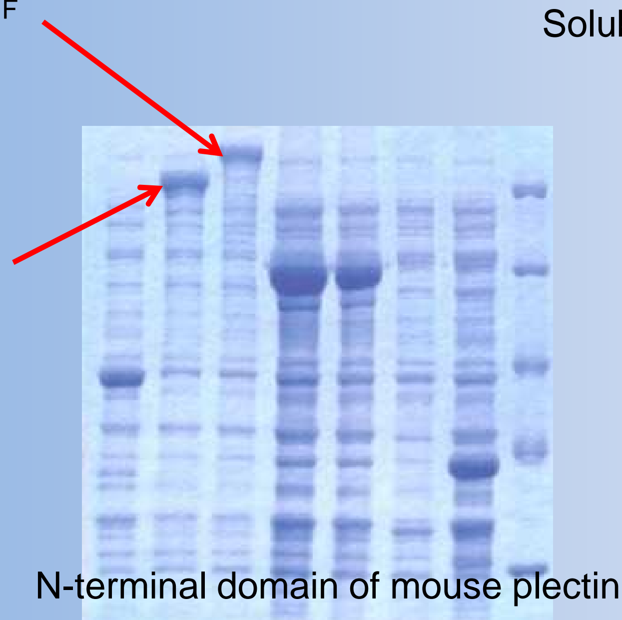
5.3.1. *E. coli* expression system



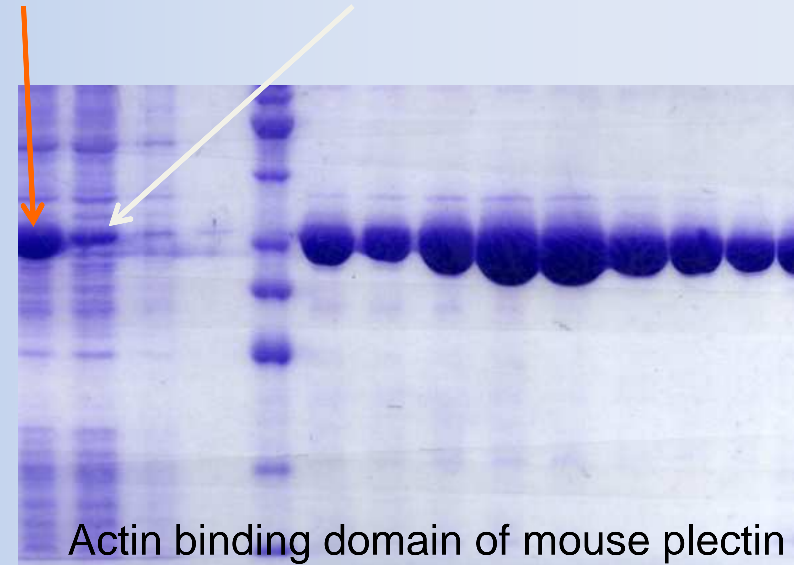
1c-30
MEFHMSGEDS ... TLRRMEEEEF
:
pI/Mw: 6.37 / **160,602.22**

1c-24
MEFHMSGEDS ...
CISELKDIEF :
pI/Mw: 6.38 / **119,799.86**

Soluble form Inclusion bodies



N-terminal domain of mouse plectin



Actin binding domain of mouse plectin

5.3. Expression system

5.3.2. Baculovirus protein expression system

- HT-bacmid propagation
- HT-suspension-based insect cell transfection
- Methods of recombinant viral titer determination

  GFP co-expression, titration assay using Alamarblue, Cedex cell counter
- HT-miniaturized deep-well block insect cell expression
- Transient insect cell expression

5.3. Expression system

5.3.2. Baculovirus protein expression system

Overview on commercially available baculovirus expression systems

Baculovirus expression kits and vendors	Compatible transfer vectors	Methodology for cloning foreign gene into transfer vector	Transfer of foreign gene into Baculovirus genome	Selection/ Recombination efficiency
BacPAK™ (Clontech)	Based on homologous recombination at polyhedrin locus	Ligase dependent	Homologous recombination in insect cells	≥90%
Bac-to-Bac™ (Invitrogen)	Based on site-specific transposition	Ligase-dependent, Gateway™ adapted	Site-specific transposition in bacterial cells	Selection of recombinants by blue-white selection on agar plates
BaculoDirect™ (Invitrogen)	Based on site-specific recombination	Gateway™ adapted	Site-specific recombination in Eppendorf tube	Antibiotic selection of transfectants in insect cells
flashBac™ (OET/ NextGen Sciences)	Based on homologous recombination at polyhedrin locus	Ligase dependent	Homologous recombination in insect cells	100%
BacVector™ 1000, 2000, 3000 (EMD/ Novagen)	Based on homologous recombination at polyhedrin locus	Ligase dependent	Homologous recombination in insect cells	≥ 95%
BaculoGold™ (Cloentech)	Based on homologous recombination at polyhedrin locus	Ligase dependent	Homologous recombination in insect cells	≥ 95%
DiamondBac™ (Sigma-Aldrich)	Based on homologous recombination at polyhedrin locus	Ligase dependent	Homologous recombination in insect cells	≥ 95%

5.3. Expression system

5.3.3. Cell-free protein expression system

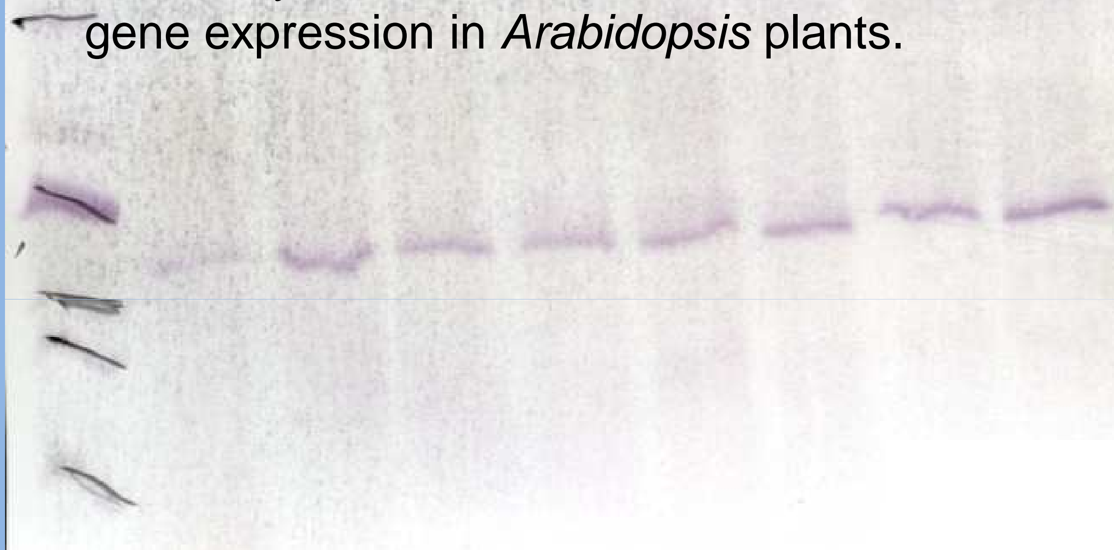
Simple open system which influences:

- Protein folding
- Disulfide bond formation
- Incorporation of unnatural amino acids
- Protein stability
- Expression of toxic proteins

Use the machinery of *E. coli* S30

5.3. Expression system

5.3.4. Transient protein expression in tobacco leaves



An *Agrobacterium*-mediated transient expression assay has been described for in vivo analysis of constitutive or inducible gene expression in *Arabidopsis* plants.

- Plant number: ca 30
- Weight of tobacco leaves: 7–10 g
- Number of tobacco leaves: 12–15
- Total: ca 3.5 kg~12–15 g protein~120–150 mg scFv