

Kód předmětu: C8980



MASARYKOVA UNIVERZITA

Protein expression and purification

## VIII. Fusion proteins and affinity purification

Lubomír Janda, Blanka Pekárová, Radka Dopitová, Lukáš Žídek

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



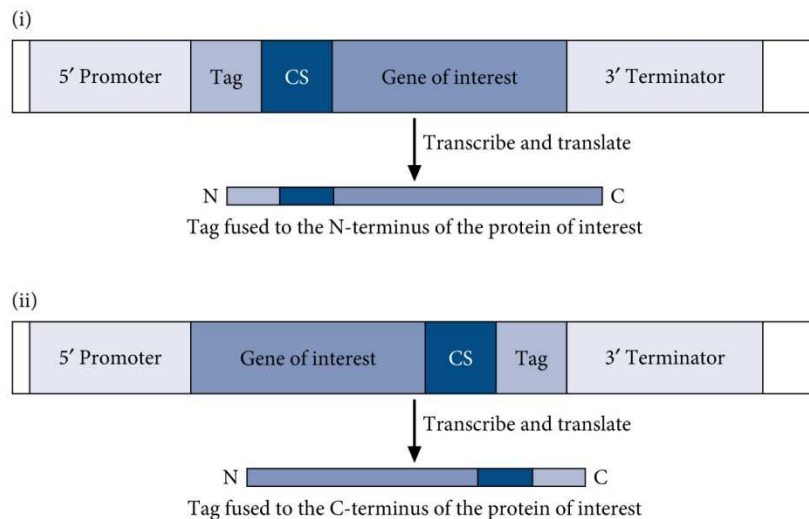
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

## Fusion proteins (tagged proteins)

Translation fusion of sequences coding a recombinant protein and

a) **short peptides** [ex. (His)<sub>n</sub>, (Asp)<sub>n</sub>, (Arg)<sub>n</sub> ... ].

b) **protein domains, entire proteins** [ex. MBP, GST, thioredoxin ...].



Engineering a tagged protein requires adding the DNA encoding the tag to either the **5' or 3' end** of the gene encoding the protein of interest to generate a single, recombinant protein with a tag at the **N- or C-terminus**. The stretch of amino acids containing a **target cleavage sequence (CS)** is included to allow selective removal of the tag.

**Expression plasmids containing various tags are commercially available.**

## Purposes of fusion tags

- **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.
- **Improving detection** – Fusion of the target protein to either terminus of a short peptide (epitope tag) or protein which is recognized by an antibody (Western blot analysis) or by biophysical methods (e.g. GFP by fluorescence) facilitates the detection of the resulting protein during expression or purification.
- **Localization** – A tag, usually located on the N-terminus of the target protein, which acts as an address for sending a protein to a specific cellular compartment.
- **Facilitating the purification of recombinant proteins** – Simple purification schemes have been developed for proteins used at either terminus which bind specifically to affinity resins.

**No single tag is ideally suited for all of these purposes.**

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

# Combinatorial tagging

➤ **No single tag is ideally suited for all purposes. Therefore, combinatorial tagging might be the only way to harness the full potential of tags in a high-throughput setting.**

## **Combinations:**

**Solubility-enhancing tag + purification tag:** MBP + His<sub>6</sub> tag

**2x purification tag:** IgG-binding domain + streptavidin-binding domain

**Localization tag + purification tag:** GFP + His<sub>6</sub> tag

**Localization tag + 2x purification tag + immunodetection:** GFP + SBP domain + His<sub>8</sub> tag + c-Myc

## Advantages and disadvantages of used fusion tags

Tag <sup>a</sup>	Advantages	Disadvantages
GST	Efficient translation initiation Inexpensive affinity resin Mild elution conditions	High metabolic burden Homodimeric protein Does not enhance solubility
MBP	Efficient translation initiation Inexpensive affinity resin Enhances solubility Mild elution conditions	High metabolic burden
NusA	Efficient translation initiation Enhances solubility Not an affinity tag	High metabolic burden
Thioredoxin	Efficient translation initiation Enhances solubility	Not an affinity tag <sup>b</sup>
Ubiquitin	Efficient translation initiation Might enhance solubility	Not an affinity tag
FLAG	Low metabolic burden High specificity	Expensive affinity resin Harsh elution conditions
BAP	Low metabolic burden Mild elution conditions  Provides convenient means of immobilizing proteins in a directed orientation	Expensive affinity resin Variable efficiency of enzymatic biotinylation Co-purification of <i>E. coli</i> biotin carboxyl carrier protein on affinity resin Does not enhance solubility
His <sub>6</sub>	Low metabolic burden  Inexpensive affinity resin Mild elution conditions Tag works under both native and denaturing conditions	Specificity of IMAC is not as high as other affinity methods  Does not enhance solubility
STREP	Low metabolic burden High specificity	Expensive affinity resin Does not enhance solubility
SET	Mild elution conditions Enhances solubility	Not an affinity tag
CBP	Low metabolic burden High specificity	Expensive affinity resin Does not enhance solubility
S-tag	Mild elution conditions Low metabolic burden High specificity	Expensive affinity resin Harsh elution conditions (or on-column cleavage) Does not enhance solubility

<sup>a</sup>GST, glutathione S-transferase; MBP, maltose-binding protein; NusA, N-utilization substance A; FLAG, FLAG-tag peptide; BAP, biotin acceptor peptide; His<sub>6</sub>, hexahistidine tag; STREP, streptavidin-binding peptide; SET, solubility-enhancing tag; CBP, calmodulin-binding peptide.

<sup>b</sup>Derivatives of thioredoxin have been engineered to have affinity for immobilized metal ions (His-patch thioredoxin) or avidin/streptavidin [38].

Waugh, 2005

➤ **Proteins do not naturally lend themselves to high-throughput analysis because of their diverse physiological properties. Affinity tags have become indispensable tools for structural and functional proteomics.**

**X**

➤ **Because affinity tags have the potential to interfere with structural and functional studies, provisions must also be made for removing them.**

# Increasing the yield of recombinant proteins using fusion technology

**Yield enhancing tags are proteins and peptides which can be involved in:**

➤ **Increasing the efficiency of translation initiation (e.g. GST, MBP, NusA...)**

- Advantage of N-terminal tags
- Providing a reliable context for efficient translation initiation
- Ribosome efficiently initiates translation at the N-terminal methionin of the tag
- Deleterious secondary structures are more likely to occur in conjunction with short N-terminal tags because short RNA-RNA interactions tend to be more stable than long-range interactions.

➤ **Protection against proteolytic degradation**

- Several studies have shown that the nature of terminal residues in a protein can play a role in recognition and subsequent action by proteases and in some cases affinity tags might improve the yield of recombinant proteins by rendering them more resistant to intracellular proteolysis.

➤ **Helping to properly fold their partners leading to increased solubility of the target protein (*in vivo* and *in vitro*).**

# Enhancing the solubility of recombinant proteins

## Solubility-enhancing tags

- Fusion of a soluble fusion partner often improves the solubility (proper folding) of the target protein.
- Advantage of N-terminal tags
- Rather proteins (highly soluble proteins) than peptides
- They are not universal
- The mechanism by which partners exert their solubilising function is not fully understood.

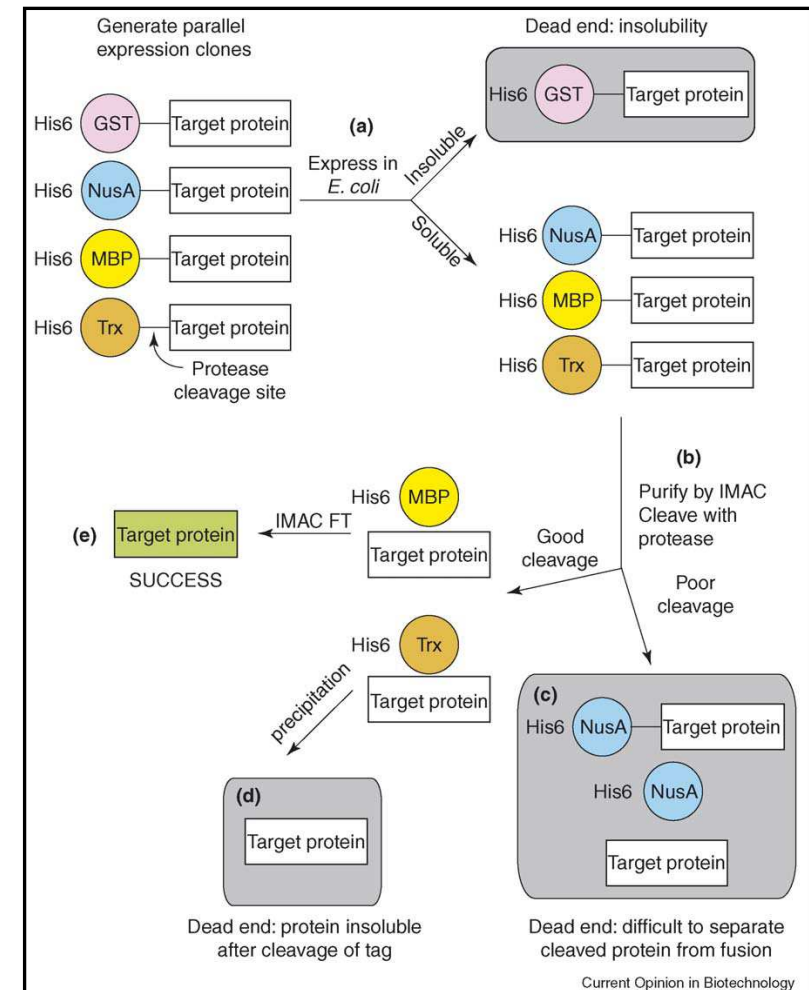
## ➤ PROTEINS

Some commonly used solubility-enhancing fusion partners		
Tag	Protein	Source organism
MBP	Maltose-binding protein	<i>Escherichia coli</i>
GST	Glutathione-S-transferase	<i>Schistosoma japonicum</i>
Trx	Thioredoxin	<i>Escherichia coli</i>
NusA	N-Utilization substance	<i>Escherichia coli</i>
SUMO	Small ubiquitin-modifier	<i>Homo sapiens</i>
SET	Solubility-enhancing tag	Synthetic
DsbC	Disulfide bond C	<i>Escherichia coli</i>
Skp	Seventeen kilodalton protein	<i>Escherichia coli</i>
T7PK	Phage T7 protein kinase	Bacteriophage T7
GB1	Protein G B1 domain	<i>Streptococcus</i> sp.
ZZ	Protein A IgG ZZ repeat domain	<i>Staphylococcus aureus</i>

Adopted from Esposito and Chatterjee, 2006

## ➤ PEPTIDES

- Poly-Arg
- Poly-Lys

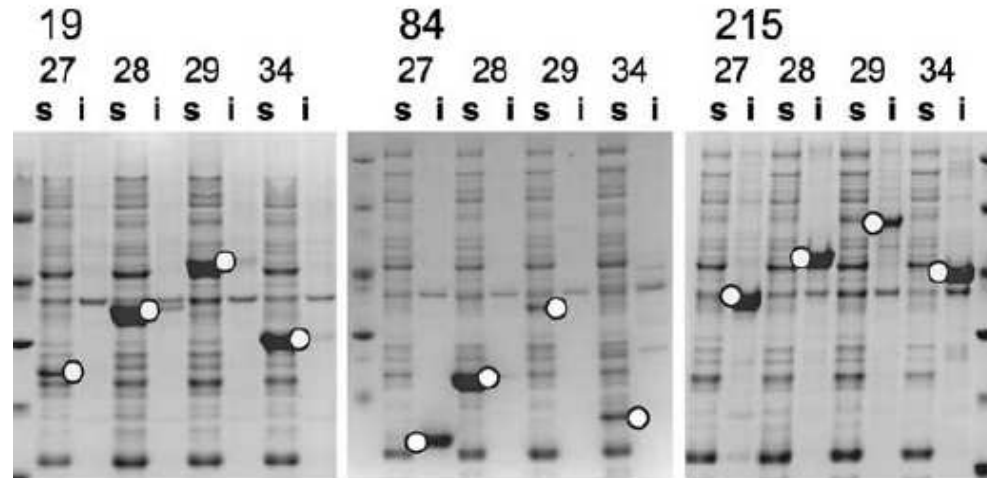


**Schematic representation of the pathway from protein expression to purification using solubility tags** (Esposito and Chatterjee, 2006).



## Enhancing the solubility of recombinant proteins

19, 84, 215 – human proteins involved in cancer produced in *E.coli*



Example of SDS PAGE with **soluble (s) and insoluble (i) fractions** following lysis. The results produced from the four different expression vectors (**27: His tag only; 28: thioredoxin + His tag; 29: GST + His tag; 34: GB1 + His tag**) are shown for three different target proteins (*Hammarstrom et al., 2006*).

## Solubility-enhancing tags - the mechanism of action

- The mechanism by which partners exert their solubilising function is not fully understood (they might act through a chaperone-like mechanism)
- possibly differs between fusion proteins

### Examples of possible mechanisms

**Maltose binding protein** might bind reversibly to exposed hydrophobic regions of nascent target polypeptide, steering the polypeptides towards their native conformation by a chaperone like mechanism

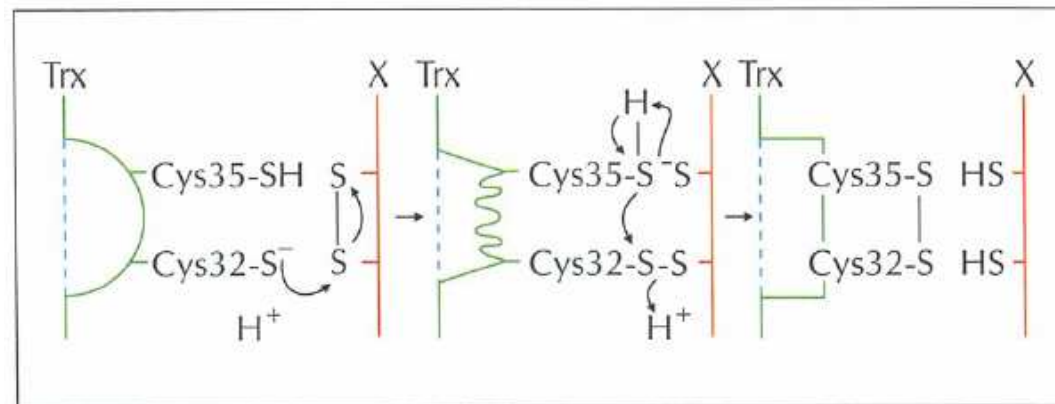
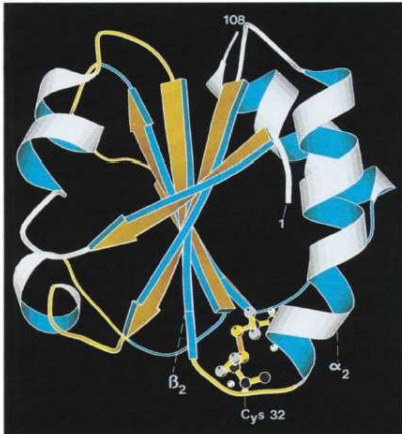
**NusA** decreased translation rates by mediating transcriptional pausing, that might enable critical folding events to occur.

**Highly acidic tags** (peptide) inhibit aggregation by increasing electrostatic repulsion between nascent polypeptides (Zhang et. 2004) .

# Solubility-enhancing tag – mechanism of action

## Thioredoxin

- Serves as a general protein disulfide oxidoreductase.
- The active-site surface in thioredoxin is designed to fit many proteins. Thioredoxin thus uses a chaperone-like mechanism of conformational changes to bind a diverse group of proteins and fast thiol-disulfide exchange chemistry in a hydrophobic environment to promote high rates of disulfide reduction.
- Increasing the solubility of the target proteins by overproduction of thioredoxin strongly suggests that the redox state affects the solubility of target proteins.



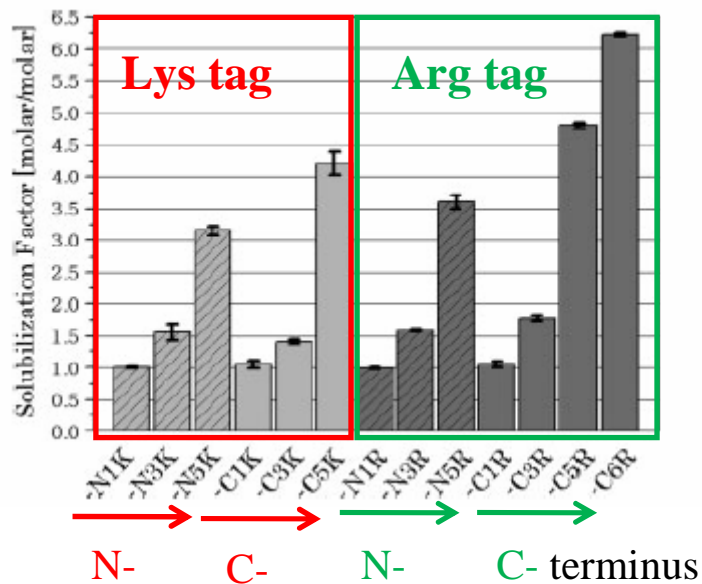
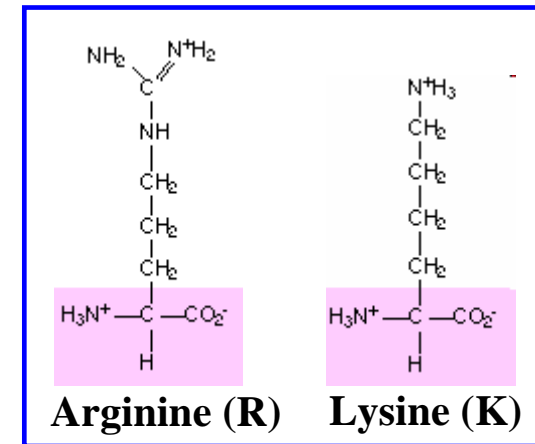
Proposed mechanism of thioredoxin-catalyzed protein disulfide reduction. Reduced thioredoxin [Trx-(SH)<sub>2</sub>] binds to a target protein via its hydrophobic surface area. Nucleophilic attack by the thiolate of Cys32 results in formation of a transient mixed disulfide, which is followed by nucleophilic attack of the deprotonated Cys35 generating Trx-S<sub>2</sub> and the reduced protein. Conformation changes in thioredoxin and the target protein occur during the reaction.

## In vitro solubility-enhancing tags

### Short peptide tags

**Poly-Lys tag, poly-Arg tag** = one, three and five lysine or arginine residues fused to the C- or N-terminus of the target protein

*Solubility as defined here is the maximum protein concentration of the supernatant after centrifugation of the supersaturated protein sample (in vitro solubility).*



BPTI-22 = bovine pancreatic trypsin inhibitor variant containing 22 alanines

The solubilization factor is defined as the molar ratio between the solubility of tagged BPTI-22 variants and that of the reference BPT-22 molecule.

The solubilization effect of poly-Lys tags is lower than that of poly-Arg tags (lysines are less hydrophilic than arginines).

## Biochemical properties of poly-Arg and poly- Lys tagged BPTI-22 protein

Protein	Protein Solubility			Rel. Trypsin Inhibitory Activity (%) <sup>c</sup>
	Conc. [mM] (Conc. [mg/ml]) <sup>a</sup>	Solubilization Factor <sup>b</sup>	T <sub>m</sub> (°C)	
BPTI-22	1.70 (10.00)	—	38.4	—
-N1K	1.70 (10.40)	1.00 (1.04)	35.2	1.05
-N3K	2.66 (19.97)	1.56 (2.00)	34.4	1.04
-N5K	5.37 (35.60)	3.16 (3.56)	34.3	1.05
-C1K	1.79 (10.95)	1.05 (1.10)	34.6	1.05
-C3K	2.41 (15.28)	1.42 (1.53)	36.2	1.05
-C5K	7.16 (47.47)	4.21 (4.75)	35.0	1.02
-N1R	1.69 (10.34)	0.99 (1.03)	35.5	1.02
-N3R	2.70 (17.23)	1.59 (1.72)	35.6	0.99
-N5R	6.20 (41.11)	3.65 (4.11)	35.5	0.99
-C1R	1.81 (11.07)	1.06 (1.11)	35.0	1.05
-C3R	3.02 (19.26)	1.78 (1.93)	34.4	1.05
-C5R	8.23 (54.56)	4.84 (5.46)	34.8	1.08
-C6R	10.59 (73.41)	6.22 (7.34)	32.7	1.1
BPTI-22 <sup>d</sup>	5.63 (33.11)	3.31 (3.31)	ND <sup>e</sup>	1.09
BPTI-22 <sup>f</sup>	2.01 (11.82)	1.18 (1.18)	ND <sup>e</sup>	NA <sup>g</sup>

The **addition of 0.5 M Arg** barely increased its solubility, and trypsin activity was inhibited by the high arginine concentration. On the other hand, addition of **50 mM Arg+Glu** was more effective and increased protein solubility more than threefold.

Protein solubility was determined as the maximum supernatant concentration of a supersaturated protein solution at 4°C in 100 mM acetate buffer pH 4.7.

<sup>a</sup> Maximum concentrations calculated in milligrams per milliliter are indicated in parenthesis. The Mw of BPTI-22, -N1K and -C1K, -N3K and -C3K, -N5K and -C5K, -N1R and -C1R, -N3R and -C3R, -N5R and -C5R, and -C6R are, respectively: 5880, 6123, 6379, 6636, 6151, 6463, 6776, and 6932 Da.

<sup>b</sup> Calculated as the ratio between the molar protein solubility of BPTI-22 and that of tagged BPTI-22. Values in parenthesis indicate the ratio calculated in milligrams per milliliters.

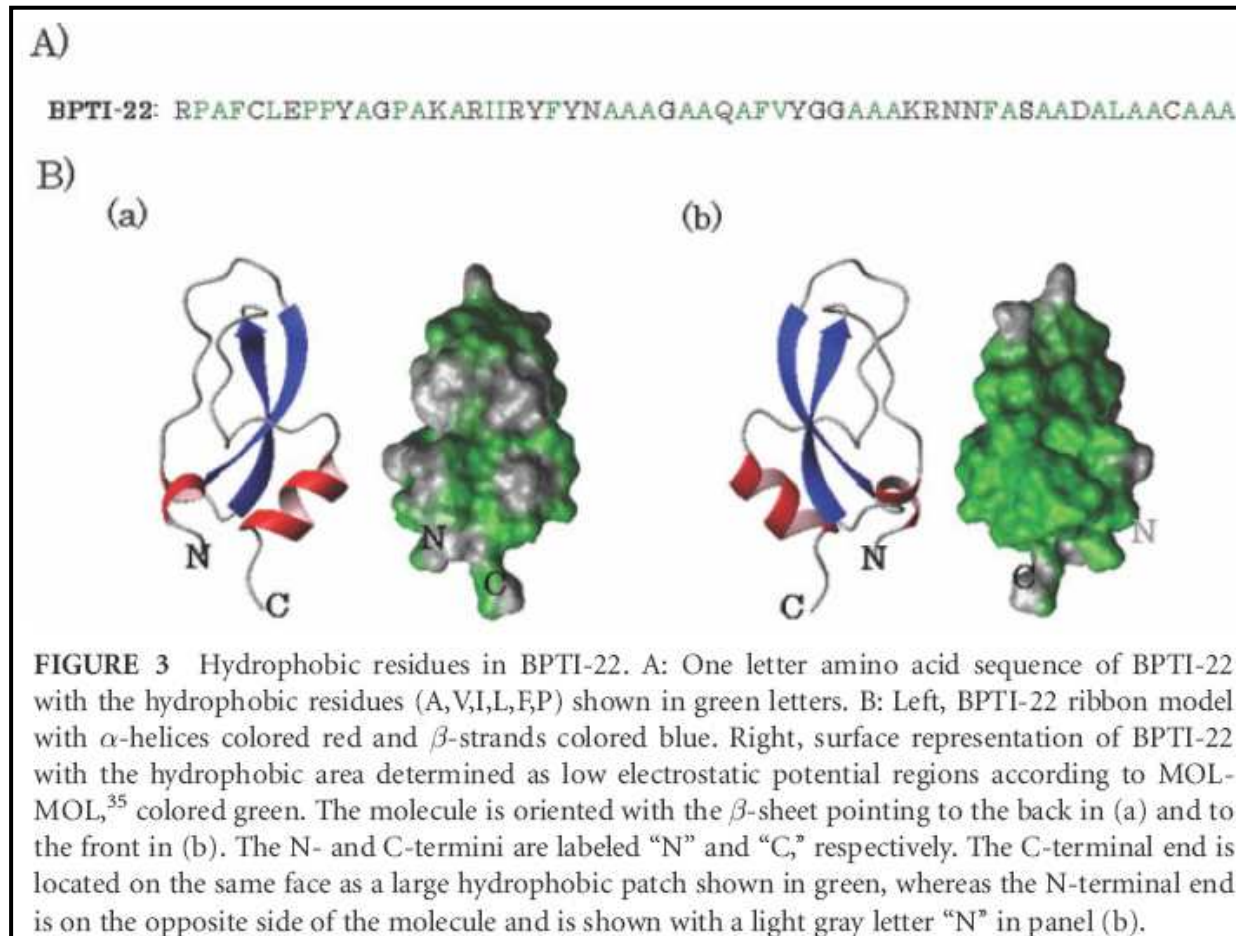
<sup>c</sup> Relative trypsin inhibitory activity calculated as the ratio between the activity of BPTI-22 and that of tagged BPTI-22. BPTI-22, which lacks R39, an arginine residue involved in two hydrogen bonding interactions with the trypsin residue backbone,<sup>34</sup> has a reduced trypsin inhibitor activity corresponding to ~60% of the wt-BPTI and BPTI-[5,55] at stoichiometry and a protein concentration of 280 nM.<sup>19</sup>

<sup>d</sup> Solubility in the same buffer as above but with the addition of 50 mM L-Arg + L-Glu.

<sup>e</sup> The CD thermal melting curve could not be determined due to the strong absorption of arginine and glutamic acid.

<sup>f</sup> Protein solubility with 500 mM Arg-HCl added to the above buffer.

<sup>g</sup> The trypsin activity could not be determined because the high arginine concentration inhibited trypsin activity.



- The solubilization factor of all C-terminal tags was slightly higher than that of the respective N-terminal tags.
- The C-terminus of BPTI-22 is close to a large hydrophobic patch, whereas the N-terminus is located on the opposite side of the molecule, away from the hydrophobic patch.
- **Charged residues seem to act through repulsive electrostatic interaction and thus hamper intermolecular interaction arising from the hydrophobic cluster.**



## Solubility-enhancing tags – comparison of peptide and protein tags, conclusions

- Protein tags are inherently large and need to be correctly folded in order to enhance solubility.
- Proteins tag are often natural affinity tags.
- Peptide tags are small, and, importantly, they do not need to be folded, which provides a significant advantage over protein tags.
- The use of small tags (< 30 amino acids long) does not increase protein size substantially and diminishes steric hindrance, which simplifies downstream structural and functional applications without the need to remove the tag.
- The solubilization enhancement effect depends on the size of the target protein. Solubility enhancement of fusion partners such as thioredoxin, GB1 is less pronounced for larger target proteins (above 25 kDa).

**MANY TAGS SUFFER FROM THE SAME PROBLEM – THEY DO NOT FUNCTION EQUALLY WELL WITH ALL TARGET PROTEINS.**

## **Removal of fusion tags- the Achilles' heel of the fusion approach**

All tags, whether small or large, have the potential to interfere with the biological activity of a protein, impede its crystallization, cause a therapeutic protein to become immunogenic or otherwise influence the target protein's behavior.

The fusion tags can be removed by:

- **Chemical cleavage**
- **Self - cleavage**
- **Enzymatic cleavage**



## Removal of fusion tags – chemical cleavage

➤ Rarely used.

**Cyanogen bromide**    Met/X

**Hydroxylamine**        Asn-Gly

1				M28V		40
		M12 M15				
MRGSHHHHHH	GMASMEKNNQ		GNGQGHNV	PN	DPNRNVDENA	
NANSAVKNNN	NEEPSDKHIK		EYLNKIQNSL	STEWSPCSVT		
CGNGIQVRIK	PGSANKPKDE		LDYANDIEKK	ICKVEKCS		

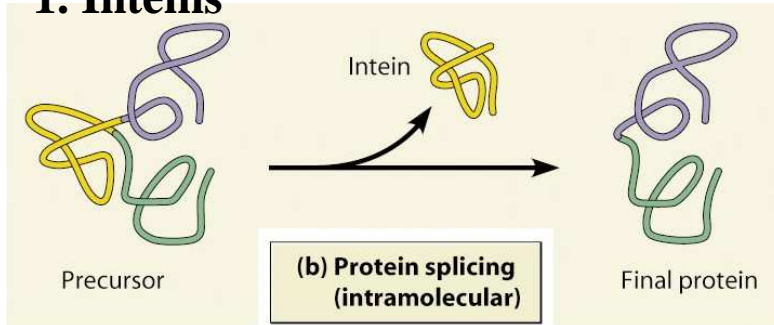
Amino – acid sequence of the *P. falciparum* C-terminal segment of CSP (PfCSP C-ter) fused to a purification tag (*Rais-Beghdadi et al., 1998*).

Chemical cleavage is a harsh method, efficient, but rather non-specific and may lead to unnecessary denaturation or modification of the target protein.

# Removal of fusion tags - Self - cleaving

## ➤ Use of self-processing fusion tags

### 1. Inteins

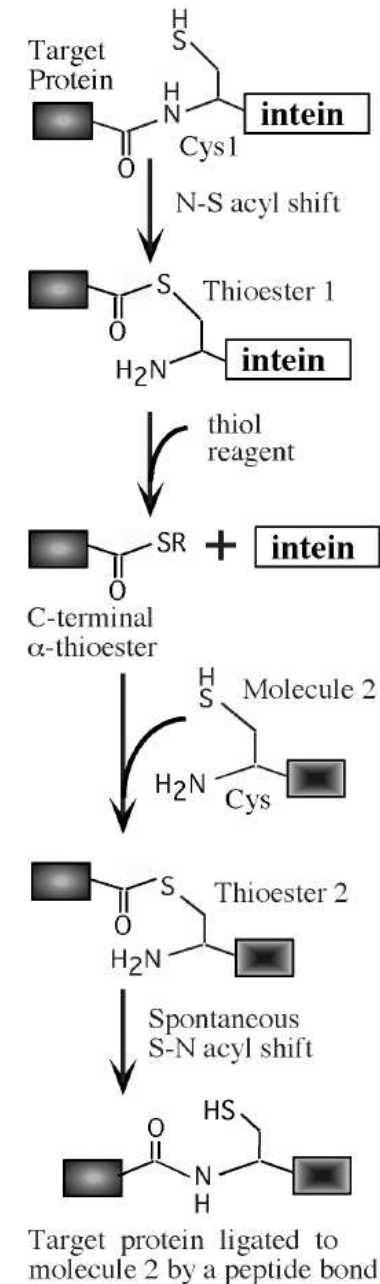
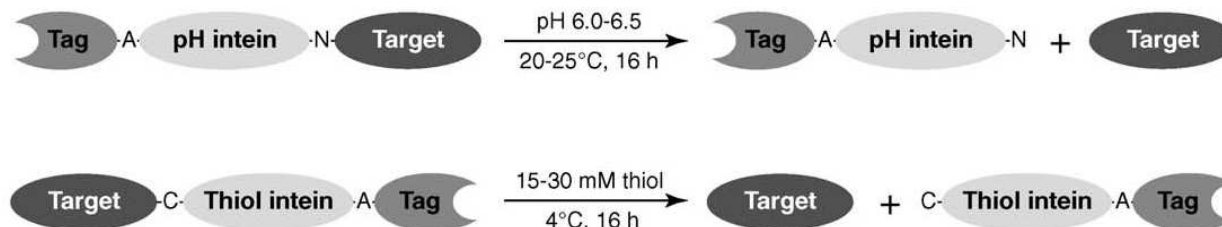


**Inteins** (*intervening proteins*) are protein segments that can excise themselves from protein precursors in which they are inserted and rejoin the flanking regions.

➤ Self - splicing inteins can be mutated at the N- or C- terminal splice junction to yield self cleaving inteins, which can be used to mediate self cleaving of various tags.

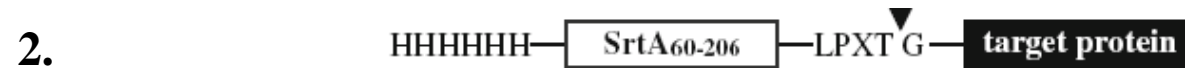
➤ Two categories of inteins:

- inteins with pH-induced C-terminal cleaving activity
- inteins with thiol-induced N- and C-terminal cleaving activity

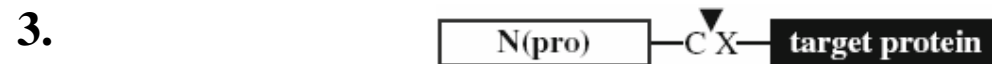


Perler, (2005)

## Removal of fusion tags - Self – cleaving fusion tag



System based on the catalytic domain of sortase A (SrtA). SrtA cleaves the Thr-Gly bond at the conserved LPXTG motif in the substrates. Cleavage is inducible by adding calcium (cofactor of SrtA).



N (pro) protein from the pestivirus polyprotein. It possesses autoproteolytic activity and catalyzes the cleavage under cosmotropic conditions.



FrpC modul: FrpC protein undergoes calcium – inducible autocatalytic processing at the peptide bond between residues Asp and Pro. Cleavage reaction is catalyzed by a self processing modul (SPM).



Vibrio cholerae secretes a large multifunctional autoprocessing repeats-in-toxin (MARTX) toxin that undergoes proteolytic cleavage during translocation into host cells. Proteolysis of the toxin is mediated by a conserved internal cysteine protease domain (CPD), which is activated upon binding of inositol hexakisphosphate.

## Removal of fusion tags - Self – cleaving fusion tag

### Inteins

- Uncontrolled in vivo cleavage or in complete in vitro cleavage
- Target protein modification – pH or thiols can modify the target protein
- Protein compatibility with cleaving conditions – pH induced inteins
- Compared to the traditional protease based method, the intein-based approach requires fewer steps and lower costs.

### Other system (2-5)

- Tested on limited number of cases

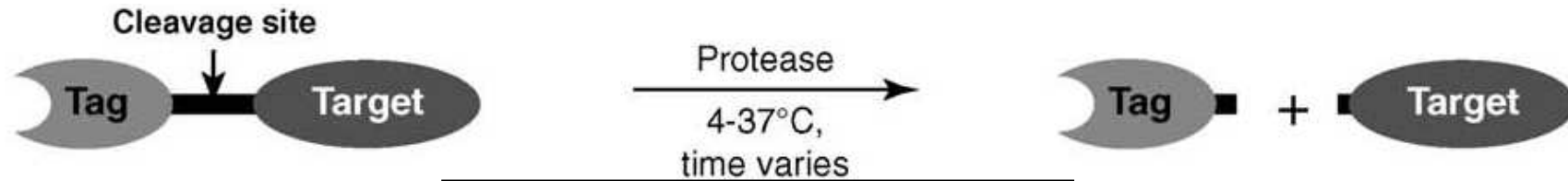
**Table 3** General features of the five self-cleavage fusion systems discussed in the text

Self-cleaving tag	MW <sup>a</sup> (kDa)	Purification tag	Cleavage condition	Advantages	Disadvantages
Intein	51; 22; 17; 15 <sup>b</sup>	CBD, CBM, phasin, ELP	Thiols; pH and/or temperature shift	Flexible fusion and cleavage options; allowing generation of target protein with native sequence	Lack of solubility-enhancing capacity; in vivo cleavage; incomplete cleavage; miscleavage
SrtA	17	His-tag, biotin	5 mM Ca <sup>2+</sup>	Potential of enhancing target protein expression and solubility	In vivo cleavage; incomplete cleavage; introduction of an extra Gly residue to the N-terminus of the target protein
N <sup>pro</sup>	19	His-tag	Kosmotropic conditions	Allowing generation of target protein with native sequence	Limited to proteins capable of refolding; in vivo cleavage; incomplete cleavage; long cleavage time
FrpC	26	His-tag, CBD	10 mM Ca <sup>2+</sup>	Efficient and tightly controlled cleavage; insensitive to protease inhibitors	Lack of solubility-enhancing capacity; introduction of an extra Asp residue to the C-terminus of the target protein; single C-terminal fusion option
CPD	23	His-tag	50–100 μM InsP6	Potential of enhancing target protein expression and solubility; efficient and tightly controlled cleavage; insensitive to protease inhibitors	Introduction of up to four non-native residues to the C-terminus of the target protein; single C-terminal fusion option

<sup>a</sup> Molecular weight of the self-cleaving tag

<sup>b</sup> Inteins with different sizes are available

## Removal of fusion tags – enzymatic cleavage



### Site-specific proteolytic cleavage:

- Exopeptidases
- Endopeptidases

### Exopeptidases (aminopeptidases and carboxypeptidases):

DAPase (TAGZyme)	Exo(di)peptidase	Cleaves N-terminal. His-tag (C-terminal) for purification and removal
<i>Aeromonas</i> aminopeptidase	Exopeptidase	Cleaves N-terminal, effective on M, L. Requires Zn
Aminopeptidase M	Exopeptidase	Cleaves N-terminal, does not cleave X-P
Carboxypeptidase A	Exopeptidase	Cleaves C-terminal. No cleavage at X-R, P
Carboxypeptidase B	Exopeptidase	Cleaves C-terminal R, K

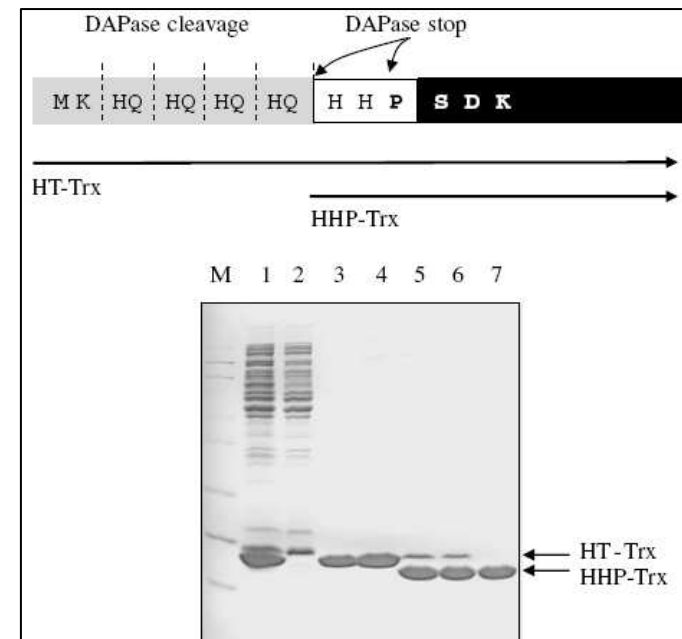
➤ APM, CPA and CPB release sequentially a single amino-acid from the N- or C- terminus of a protein until the stop site is reached.

### TAGZyme system (Qiagen):

➤ DAPase (dipeptidyl aminopeptidase I)

#### TAGZyme stop points

Amino acid	DAPase stop point (↓) sequence*
Lysine (Lys, K)	Xaa-Xaa...Xaa-Xaa ↓ Lys-Xaa ...
Arginine (Arg, R)	Xaa-Xaa Xaa-Xaa ↓ Arg-Xaa
Proline (Pro, P)	Xaa-Xaa...Xaa-Xaa ↓ Xaa-Xaa Pro-Xaa...
Proline (Pro, P)	Xaa-Xaa...Xaa-Xaa ↓ Xaa-Pro Xaa-Xaa...
Glutamine (Gln, Q)	Xaa-Xaa Xaa-Xaa ↓ Gln-Xaa..



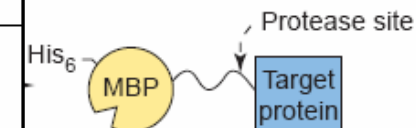
Arnau et al., 2006

# Removal of fusion tags - enzymatic cleavage

## Endopeptidases

➤ The enzymatic cleavage site has to be placed between the fusion tag and the target protein.

Enzyme	Cleavage site	Comments
Enterokinase	DDDDK*	Secondary sites at other basic aa
Factor Xa	IDGR*	Secondary sites at GR
Thrombin	LVPR*GS	Secondary sites. Biotin labeled for removal of the protease
PreScission	LEVLFQ*GP	GST tag for removal of the protease
TEV protease	EQLYFQ*G	His-tag for removal of the protease
3C protease	ETLFQ*GP	GST tag for removal of the protease
Sortase A	LPET*G	Ca <sup>2+</sup> -induction of cleavage, requires an additional affinity tag (e.g., his-tag) for on column tag removal
Granzyme B	D*X, N*X, M*N, S*X	Serine protease. Risk for unspecific cleavage



**Table 4** Cleavage (%) of enterokinase through densitometry (Hosfield and Lu 1999) based on the amino acid residue X<sub>1</sub>. The sequence...-GSDYKDDDDK-X<sub>1</sub>-ADQLTEEQIA-... of a GST-calmodulin fusion protein was tested using 5 mg protein digested with 0.2 Uof enterokinase for 16 h at 37 °C

Amino acid in position X <sub>1</sub>	Cleavage of enterokinase (%)
Alanine	88
Methionine	86
Lysine	85
Leucine	85
Asparagine	85
Phenylalanine	85
Isoleucine	84
Aspartic acid	84
Glutamic acid	80
Glutamine	79
Valine	79
Arginine	78
Threonine	78
Tyrosine	78
Histidine	76
Serine	76
Cysteine	74
Glycine	74
Tryptophan	67
Proline	61

## Enterokinase

**Asp-Asp-Asp-Asp-Lys/X**

# Removal of fusion tags - enzymatic cleavage

A critical review of the methods for cleavage of fusion proteins  
with thrombin and factor Xa

Richard J. Jenny,<sup>a,\*</sup> Kenneth G. Mann,<sup>b</sup> and Roger L. Lundblad<sup>c,d</sup>

<sup>a</sup> *Haematologic Technologies, Inc., Essex Junction, VT, USA*

<sup>b</sup> *Department of Biochemistry, University of Vermont, Burlington, VT, USA*

<sup>c</sup> *Department of Pathology, University of North Carolina, Chapel Hill, NC, USA*

<sup>d</sup> *Roger L. Lundblad, LLC, Chapel Hill, NC, USA*

Received 27 February 2003, and in revised form 7 May 2003

---

---

**Protein  
Expression  
& Purification**

---

---

The purpose of this review was to demonstrate that both thrombin and factor Xa can hydrolyze a variety of peptide bonds in proteins.

## Sequences cleaved by thrombin in polypeptide hormones

Polypeptide homones <sup>a</sup>	Sequence cleaved
Secretin	ELSLR <small>LR</small> DSA
Secretin	ELSLR <small>LR</small> (much slower than above)
Vasoactive intestine polypeptide	DNYTR <small>LR</small> K
Vasoactive intestine polypeptide	YTR <small>LR</small> KQM
Cholecystokinin	APSGR <small>V</small> SM
Cholecystokinin	VSMIK <small>N</small> LQ
Dynorphin A	RIRPK <small>L</small> KW
Somatostatin-28	AMAPR <small>E</small> RK
Somatostatin-28	NFFW <small>K</small> TFT
Gastrin releasing peptide	KMYPR <small>G</small> NH
Salmon calcitonin	QTYPR <small>T</small> NT

<sup>a</sup> The reaction mixtures contained 0.5 NIH units thrombin and 1.0 nmol peptide in 20  $\mu$ L of 50 mM  $\text{NH}_4\text{CO}_3$ , pH 8.0, at 25 °C. The conditions were designed to obtain an enzyme/substrate ratio of 1:60 (w/w).

# Accuracy of cleavage has to be precisely verified!

pRSETB::AHP2

Enterokinase cleavage site

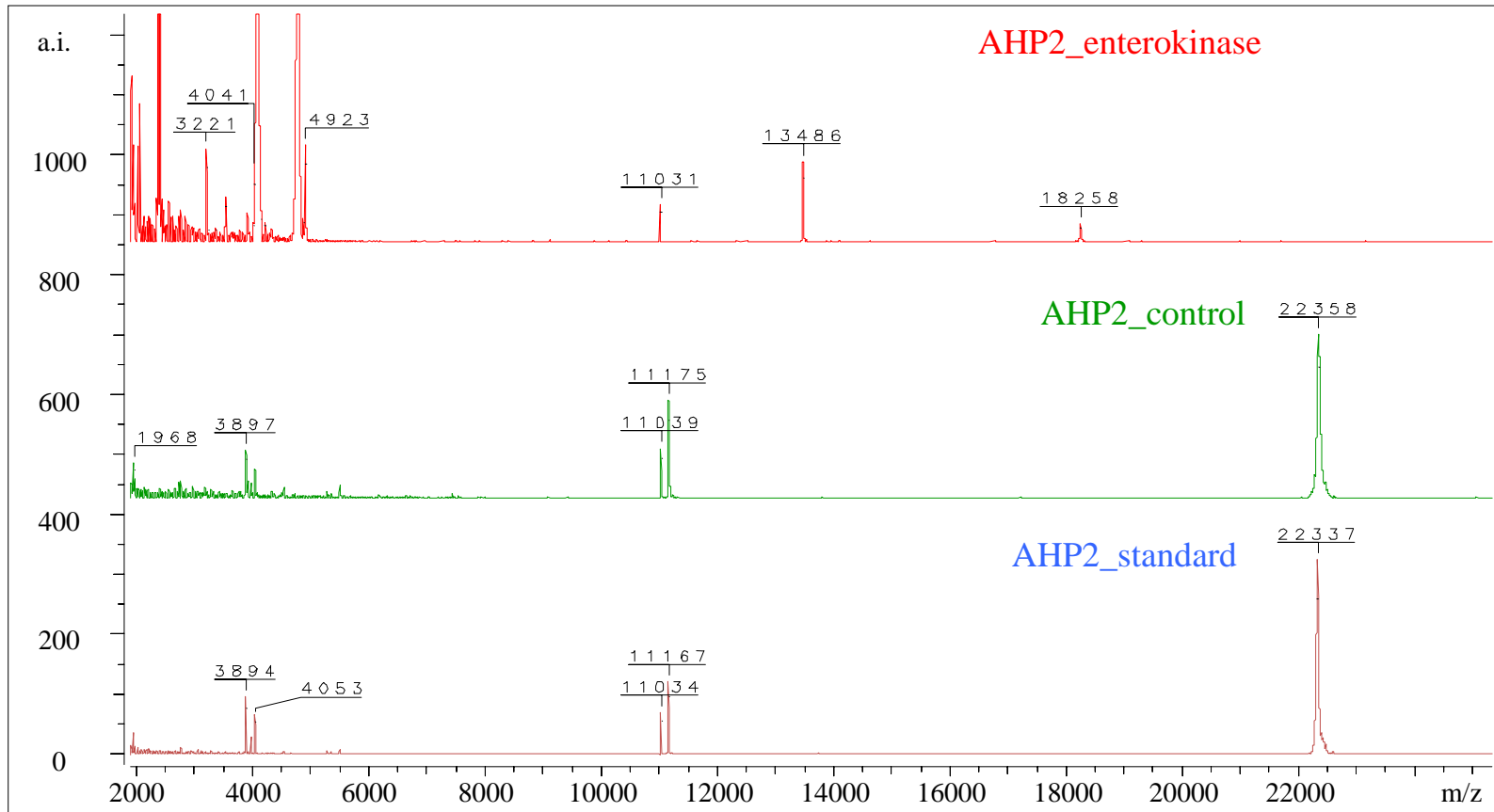
N`MRGSHHHHHHGMASMTGGQQMGRDLY**DDDDK**DPSRSAAAGTMEFMDALIA.....GIVPQVDIN C`

Theoretically:

3,4 kDa

18,9 kDa

Intact mass spectrometry analysis






## Removal of fusion tags - enzymatic cleavage

- **Unspecific cleavage** (SOLUTION: optimization of protein cleavage conditions or using re-engineered proteases with increased specificity such as ProTEV and AcTEV).
- **Optimization of protein cleavage conditions** (mainly enzyme-to-substrate ratio, temperature, pH, salt concentration, length of exposure).
- **Precipitation of the target protein when the fusion partner is removed** (so-called soluble aggregates; SOLUTION: another approach for protein solubilization has to be found).
- **Cleavage efficiency** (varies with each fusion protein in an unpredictable manner, probably due to aggregation or steric issues; the problem can be solved by introducing short linkers between the protease site and the fusion tag).
- **High cost of proteases**
- **Re-purification step**
- **Failure to recover active or structurally intact protein**
- **Target protein modification** (some proteases like thrombin, TEV, Precision leave one or two amino-acids on the target protein near the cleavage site).

### The alternative is to leave the tag in place for structural analysis:

These multi-domain proteins are usually

- less conducive to forming well-ordered diffraction crystals, presumably due to the conformational heterogeneity allowed by the flexible linker region.
- too large for NMR analysis.



The small tags are a better choice in structural biology.

# Affinity chromatography

➤ a type of adsorption chromatography, in which the molecule to be purified is specifically and reversibly adsorbed to a complementary binding substance (ligand, L) immobilized on an insoluble support (matrix, M).

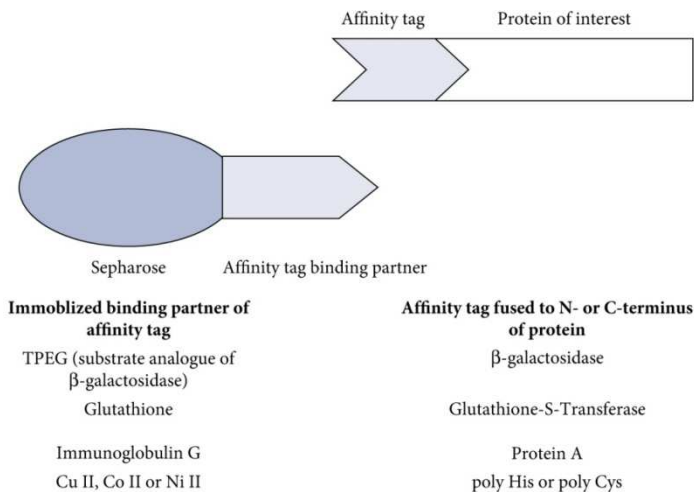
<b>Ligand</b>	<b>Affinity to</b>
<b>Enzyme</b>	<b>substrate analogue, inhibitor, cofactor</b>
<b>Antibody</b>	<b>antigen, virus, cell</b>
<b>Lectin</b>	<b>polysaccharide, glycoprotein, cell surface receptor, cell</b>
<b>Nucleic acid</b>	<b>complementary base sequence, histones, nucleic acid polymerase, nucleic acid binding protein.</b>
<b>Hormone, vitamin</b>	<b>receptor, carrier protein</b>

- AC has a concentrating effect, the high selectivities of separations derived from the natural specificities of the interacting molecules.
- AC can be used (1) to purify substances from complex biological mixtures, (2) separate native forms from denatured forms of the same substance, and (3) remove small amounts of biological material from large amounts of contaminating substances, (4) and to isolate protein complexes from the native source.
- the first application was in 1910 (adsorption of amylase onto insoluble starch) but it developed during the 1960s and 1970s.

# Affinity tags and affinity purification

**Table 2** Sequence and size of affinity tags

Tag	Residues	Sequence	Size (kDa)
Poly-Arg	5–6 (usually 5)	RRRRR	0.80
Poly-His	2–10 (usually 6)	HHHHHH	0.84
FLAG	8	DYKDDDDK	1.01
Strep-tag II	8	WSHPQFEK	1.06
c-myc	11	EQKLISEEDL	1.20
S-	15	KETAAAKFERQHMS	1.75
HAT-	19	KDHLIHNVHKEFHAAHANK	2.31
3x FLAG	22	DYKDHDGDYKDHDIDYKDDDDK	2.73
Calmodulin-binding peptide	26	KRRWKKNFIAVSAANRFKKISSGAL	2.96
Cellulose-binding domains	27–189	Domains	3.00– 20.00
SBP	38	MDEKTTGWRGGHVVEGLAGELEQLRARLEHHHPQGOREP	4.03
Chitin-binding domain	51	TNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ	5.59
Glutathione S-transferase	211	Protein	26.00
Maltose-binding protein	396	Protein	40.00



A tag is fused to the N- or C-terminus of the protein of interest to facilitate purification, which relies on a specific interaction between the affinity tag and its immobilized binding partner. **Genetically engineered fusion tags allow the purification of virtually any protein without any prior knowledge of its biochemical properties.**

# Purification tags

## Affinity tags

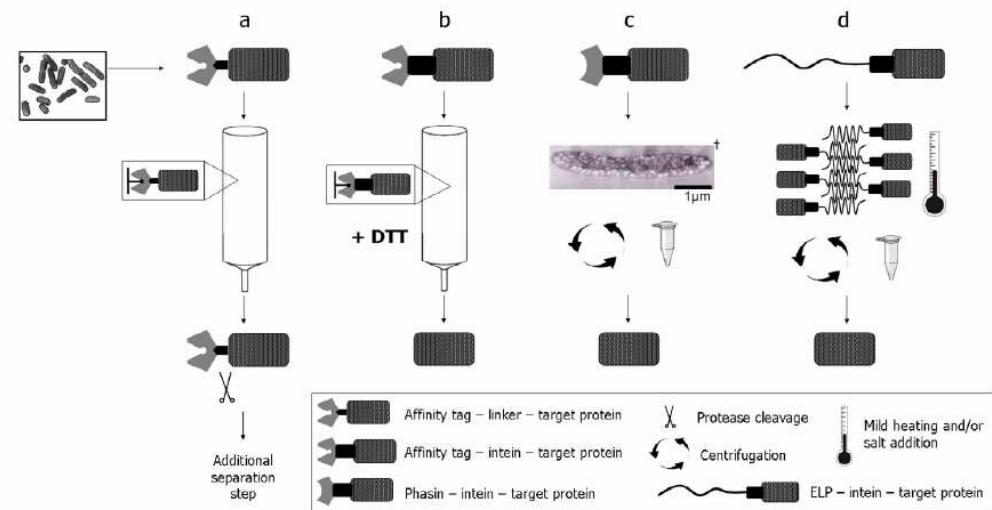
Affinity tag	Matrix
Poly-Arg	Cation-exchange resin
Poly-His	Ni <sup>2+</sup> -NTA, Co <sup>2+</sup> -CMA (Talon)
FLAG	Anti-FLAG monoclonal antibody
Strep-tag II	Strep-Tactin (modified streptavidin)
c-myc	Monoclonal antibody
S	S-fragment of RNaseA
HAT (natural histidine affinity tag)	Co <sup>2+</sup> -CMA (Talon)
Calmodulin-binding peptide	Calmodulin
Cellulose-binding domain	Cellulose
SBP	Streptavidin
Chitin-binding domain	Chitin
Glutathione S-transferase	Glutathione
Maltose-binding protein	Cross-linked amylose

## Non - chromatographic tags

Tag	Matrix
ELP	None
PHB	Intracellular PHA granules
annexin B1	None

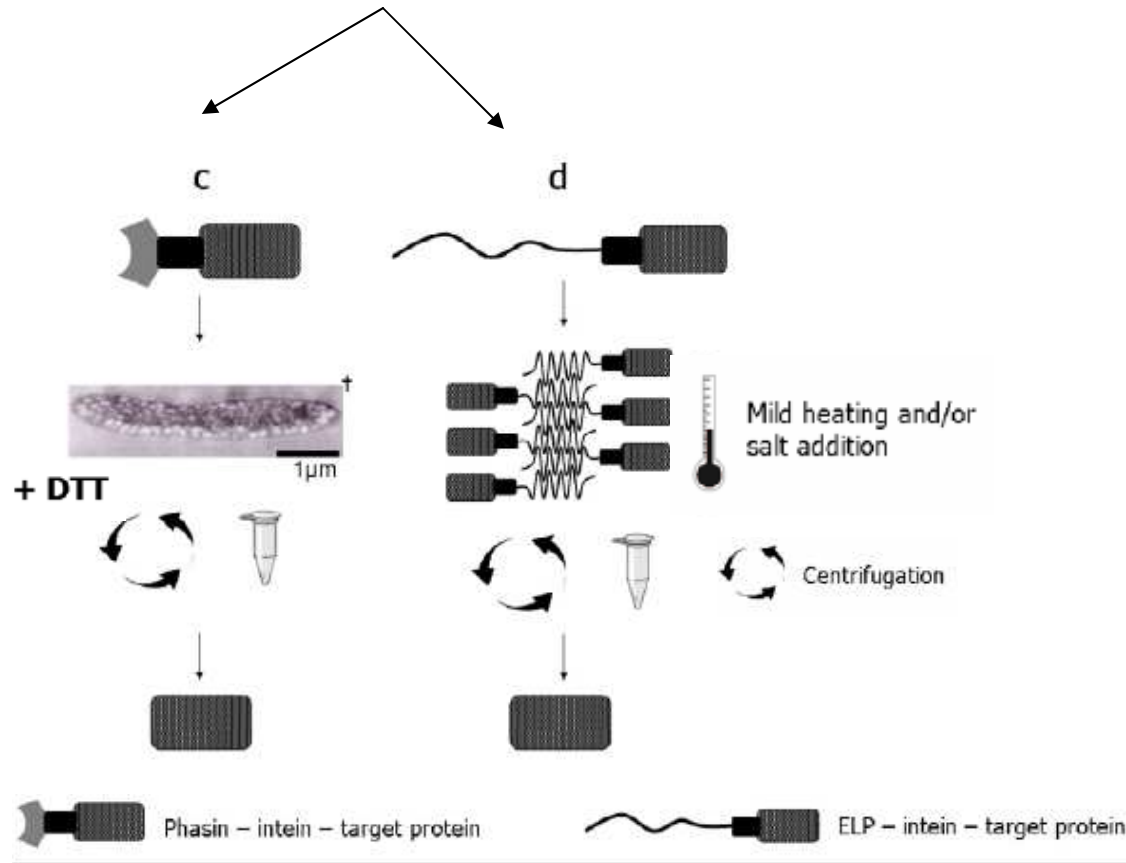
- These tags can eliminate affinity resin. Proteins are isolated by other non-chromatographic methods (centrifugation, filtration)
- typically combined with self-splicing tags
- 75 % - 95 % purity

- Traditional purification tags
- The tag binds strongly and selectively to an immobilized ligand on a solid support.
- After optimization one could achieve > 90% purity.



# Purification tags

## Non - chromatographic tags



### The PHB system (c):

➤ PHB (polyhydroxybutyrate): subclass of biodegradable polymers produced in various organisms, use as storing excess carbon.

➤ The system includes *in vivo* production of PHB small granules (from the plasmid carrying PHB-synthesis genes).

➤ Target protein in fusion to self cleaving phasin tag.

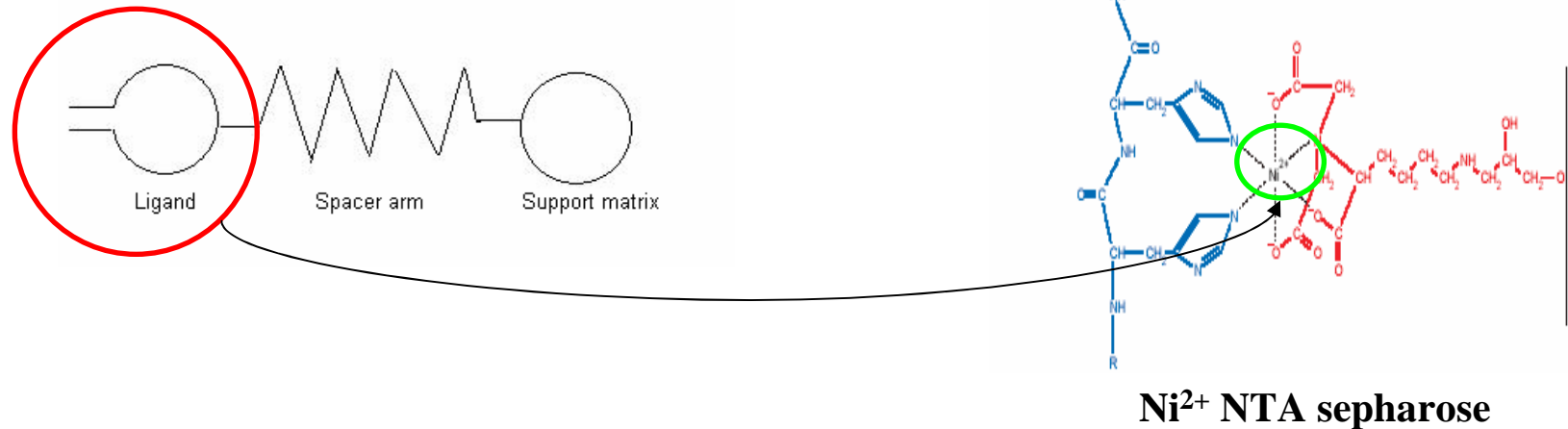
➤ Tagged protein binds to the PHB particles via phasin tag, which allows the granules and the tagged protein to be co-purified via centrifugation.

➤ DTT induced cleaving activity of intein and thus elution of the target protein.

### The ELP system (d):

➤ ELP (elastin-like polypeptide) selectively and reversibly precipitates in response to changes in temperature and buffer salts. This allows soluble and insoluble contaminants to be removed by filtration or centrifugation.

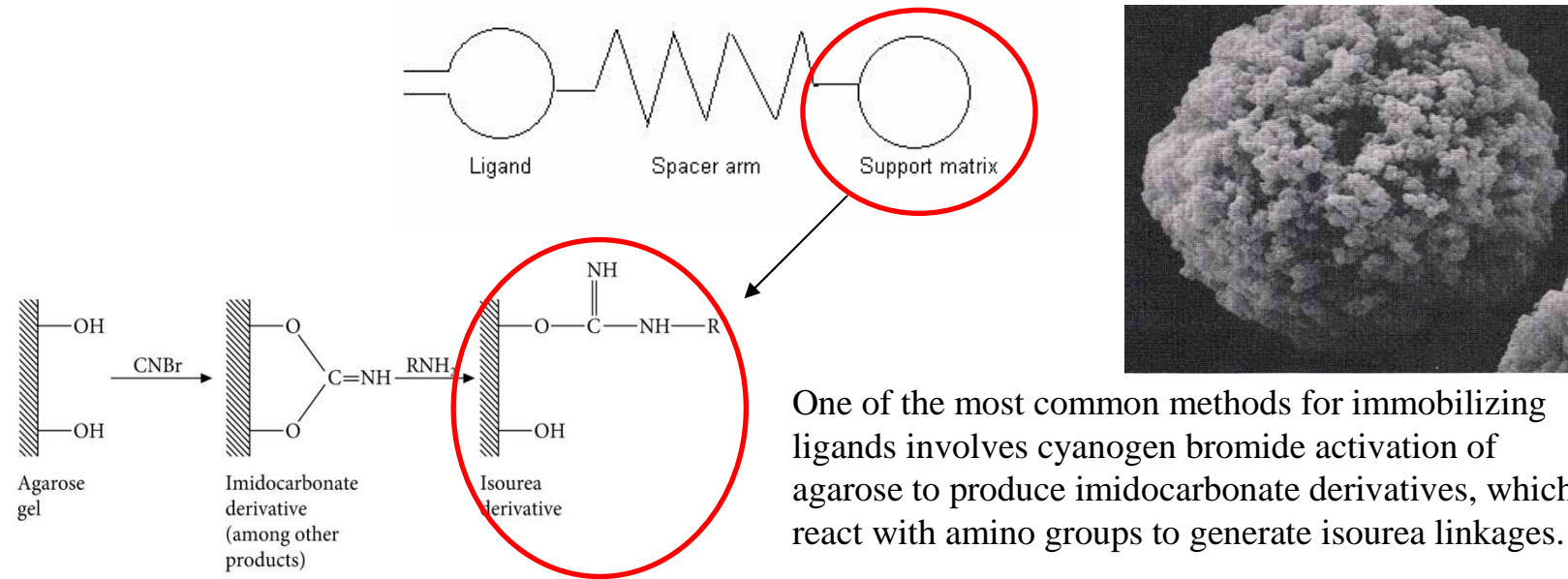
## Components of a matrix for affinity chromatography



### A ligand

- The dissociation constant ( $K_D$ ) for the ligand - target complex should ideally be in the range  $10^{-4}$  to  $10^{-8}$  M in free solution to allow efficient elution under conditions which will maintain protein stability.
- A ligand has to be attached to the matrix with a suitable chemically reactive group. The mode of attachment must not compromise the reversible interaction between the ligand and protein.

# Components of a matrix for affinity chromatography



One of the most common methods for immobilizing ligands involves cyanogen bromide activation of agarose to produce imidocarbonate derivatives, which react with amino groups to generate isourea linkages.

## A matrix

- Typically, a macroporous polysaccharide bead such as agarose, that provides a porous structure so that there is an increased surface area to which the target molecule can bind.
- A matrix has a suitable attachment site for the ligand. Typically matrices are chemically activated to permit the coupling of the ligand. A number of activation methods are available which depend on the nature of the matrix and the availability of compatible reactive groups on the ligand.

## Components of a matrix for affinity chromatography

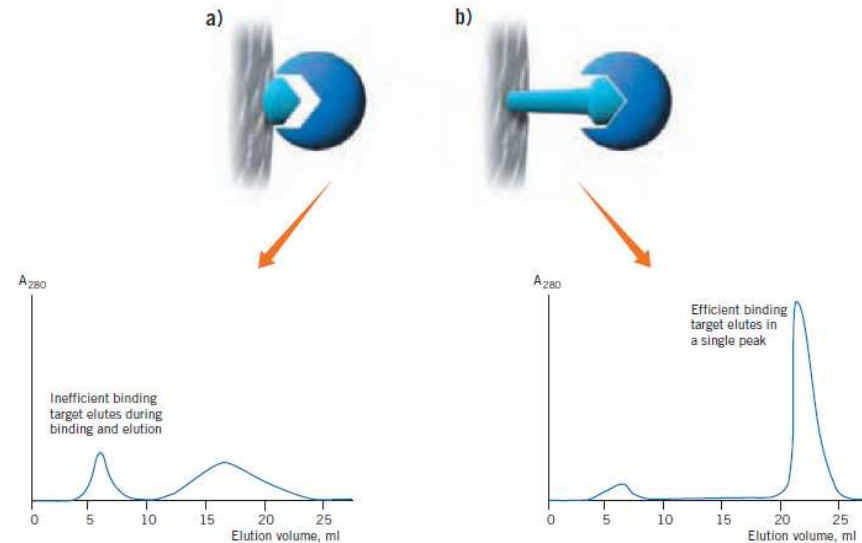
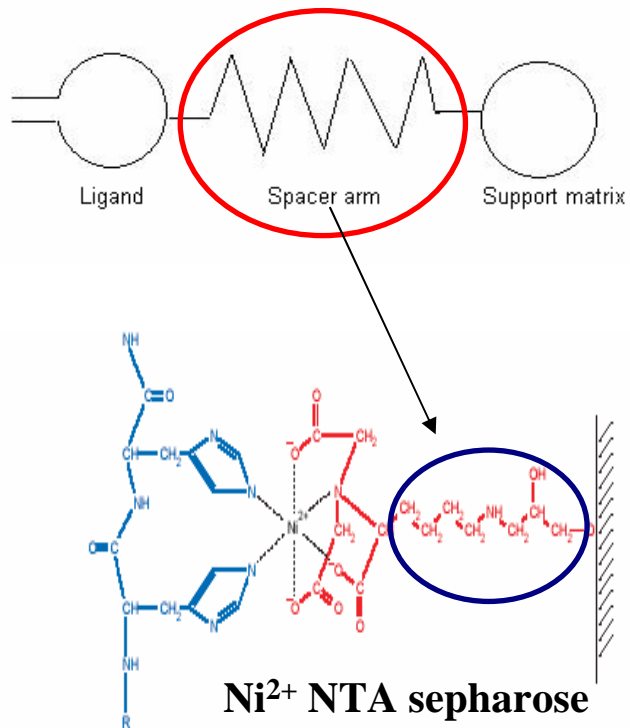


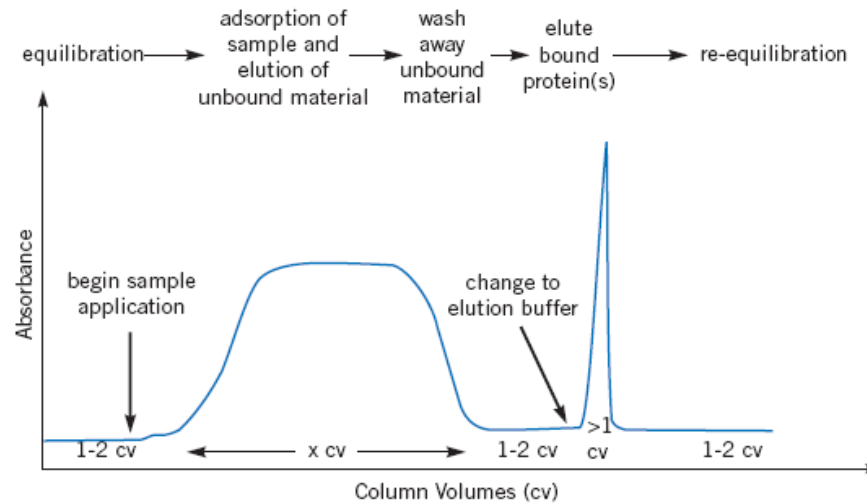
Fig. 56. Using spacer arms. a) Ligand attached directly to the matrix. b) Ligand attached to the matrix via a spacer arm.

### Spacer arm

➤ A spacer arm will be required in cases where direct coupling of the ligand to the matrix results in steric hindrance and subsequently the target protein will fail to bind to the immobilized ligand efficiently. The introduction of a spacer arm between the ligand and the matrix minimizes this steric effect and promotes optimal adsorption of the target protein to the immobilized ligand.



## Typical affinity purification steps



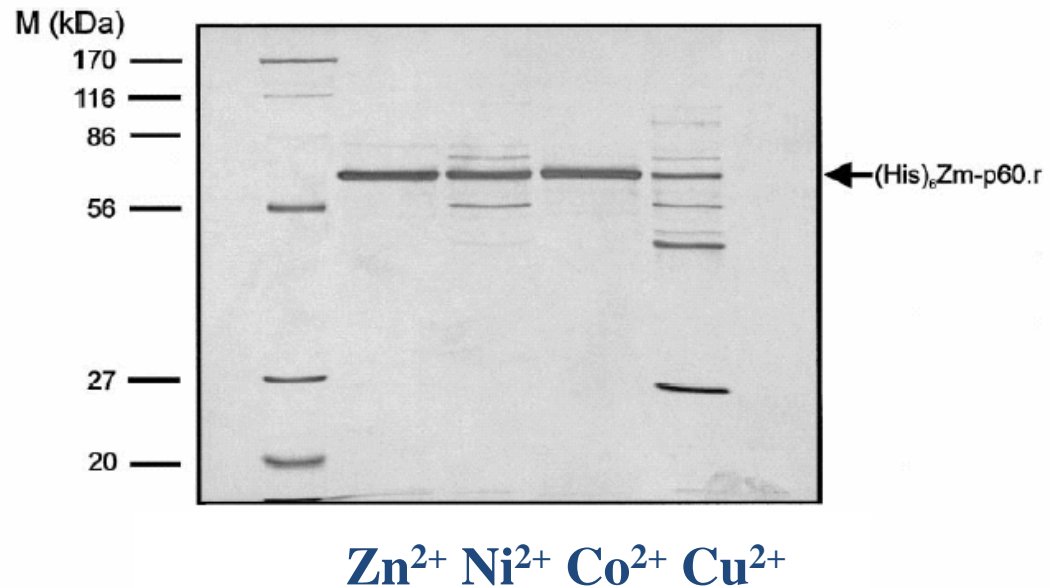
- In the equilibration phase, buffer conditions are optimized to ensure that the target molecules interact effectively with the ligand and are retained by the affinity medium as all other molecules wash through the column.
- During the washing step, buffer conditions are created that wash unbound substances from the column without eluting the target molecules or that re-equilibrate the column back to the starting conditions (in most cases the binding buffer is used as a wash buffer).
- In the elution step, buffer conditions are changed to reverse (weaken) the interaction between the target molecules and the ligand so that the target molecules can be eluted from the column.

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

- The most common purification tag is typically composed of six consecutive histidine residues.
- Histidine, cysteine, and tryptophan residues are known to interact specifically with divalent transient metal ions such as  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Zn}^{2+}$ .
- Histidine is the amino acid that exhibits the strongest interaction with immobilized metal ion matrices as the electron donor groups on the histidine imidazole ring readily form coordination bonds with an immobilized transition metal.

### Binding strength of His tag to metal ions:

$\text{Cu}^{2+} > \text{Ni}^{2+} > \text{Zn}^{2+} \sim \text{Co}^{2+}$

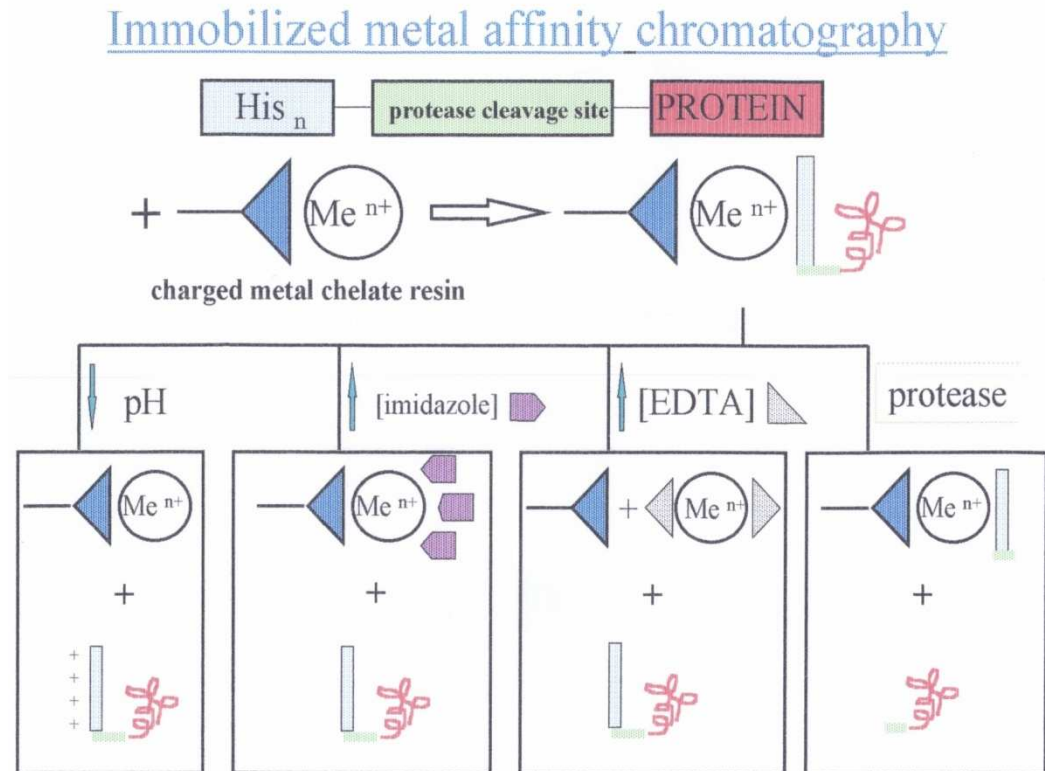


(Zouhar et al., 1999)

- IMAC can be used under native and/or denatured conditions.
- A highly purified protein can often be obtained in one or, at most, two purification steps.

## His-tagged protein and IMAC under native conditions

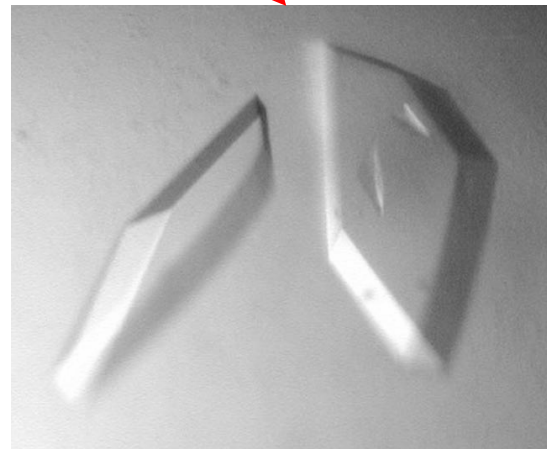
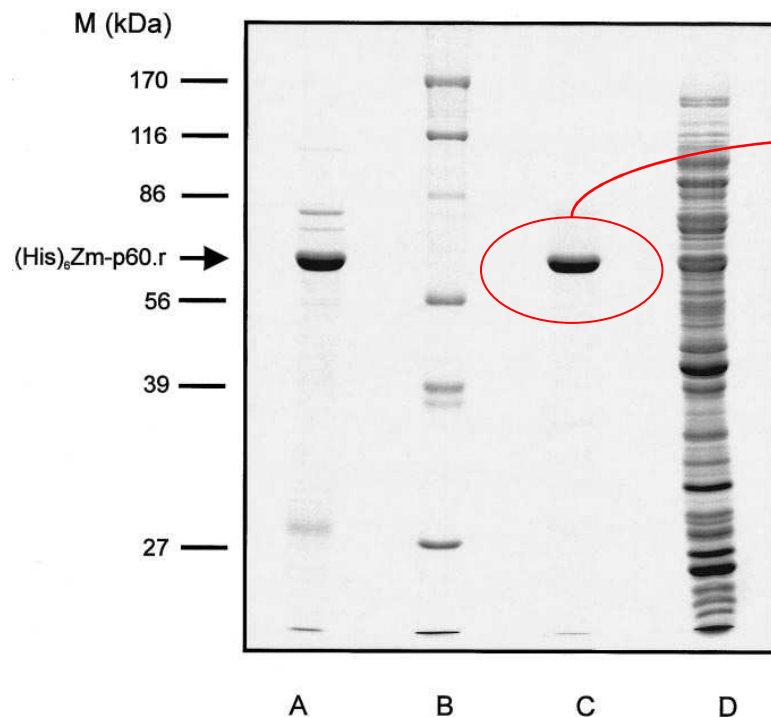
- Optimal binding of recombinant protein with metal ion is achieved at pH 7–8.
- Buffers with a high salt concentration (0.5–1 M NaCl) reduce nonspecific electrostatic interaction.
- Nonionic detergents or glycerol reduce nonspecific hydrophobic interactions.
- Elution of contaminating proteins can be 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.



## 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  $\text{Zn}^{2+}$
- Removing contaminated proteins: linear gradient of imidazole (0–50 mM) and pH (pH 7–6.1)
- Protein elution: 0.1 M EDTA
- 80% recovery, 95 fold purification
- Common production and isolation of the wild type protein and soluble mutant form for enzymatic measurements and crystallization.



(Zouhar et al., 1999)

## His-tagged protein and IMAC under denatured conditions

- Purification of proteins expressed in inclusion bodies.
- Purification in a high concentration of urea or guanidine chloride.
- Result is a pure protein, but in a denatured form (sufficient for immunization).

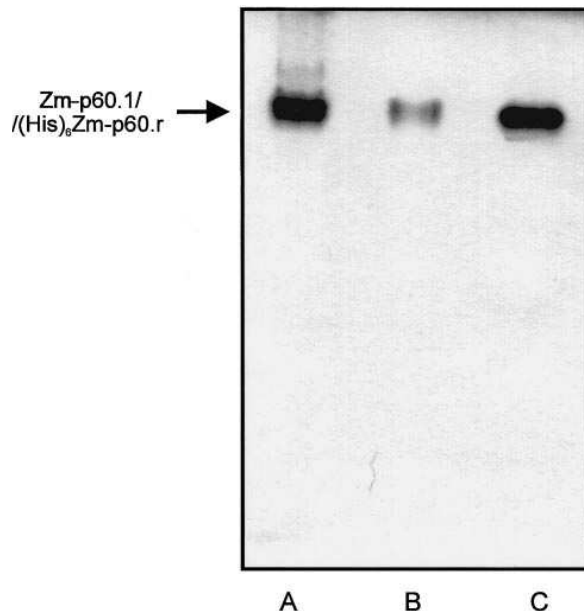
### Recovery of native conformers (necessary for functional and structural analysis):

➤ Binding to the column under strong denaturing conditions (8 M urea)

➤ **Two possibilities of renaturation:**

1. The protein is eluted from the column and renatured by dialysis or rapid dilution in renaturing buffers.
2. Renaturation of the protein bounded to the column (matrix assisted refolding procedure): gradient from denatured to renatured buffers or pulsion renaturation (8-0M urea).

Identification of properly refolded  $(\text{His})_6\text{Zm-p60.1}$  (maize  $\beta$ -glucosidase) using 10% native PAGE, followed by activity in gel staining:



**A = crude protein extract prepared from maize seedlings containing the native enzyme**

**B =  $(\text{His})_6\text{Zm-p60.1}$ , renatured product (matrix assisted refolding procedure – 23 renaturing cycles)**

**C =  $(\text{His})_6\text{Zm-p60.1}$  purified by native IMAC**

$K_M$   $(\text{His})_6\text{Zm-p60.1}$  purified by native IMAC:  **$0.64 \pm 0.06$  mM**

$K_M$   $(\text{His})_6\text{Zm-p60.1}$  renatured product:  **$0.6 \pm 0.08$  mM**

Determination of  $v_{\max}$  and  $k_{\text{cat}}$  was hampered by the fact that the refolding process yielded a number of improperly folded polypeptides.

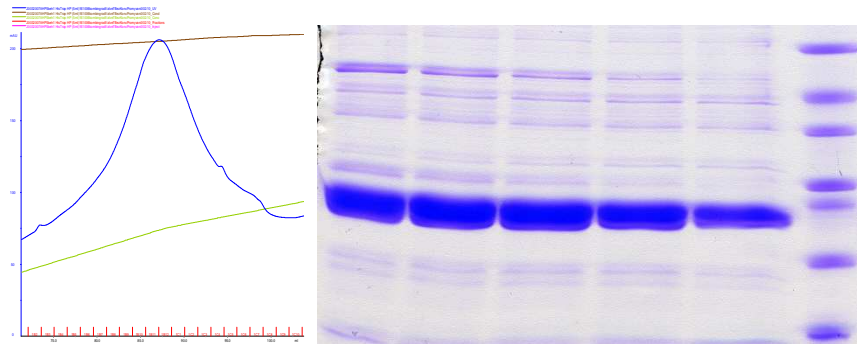
# His-tagged protein and IMAC under native conditions

## Two-step purification of *Arabidopsis* histidine phosphotransfer protein 5

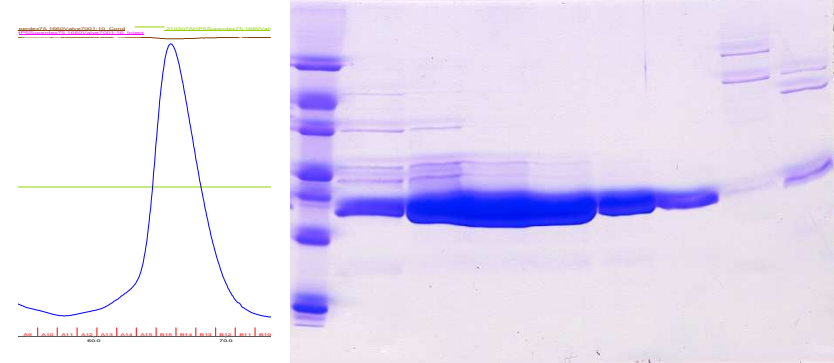
(His)<sub>6</sub>AHP5

- IMAC matrix: highly cross-linked spherical agarose
- Functional group: nitrilacetic acid , metal ion Ni<sup>2+</sup>
- Removing contaminated proteins: linear gradient of imidazole (20–500 mM)
- Protein elution: 130 mM imidazol
- Common production and isolation of the wild type protein for protein-protein interaction measurements and crystallization.

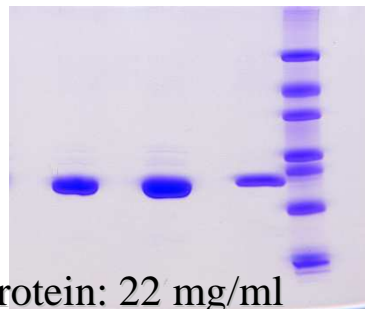
1st step - metal chelate affinity chromatography



2nd step - gel filtration



After  
ultrafiltration



Purity: 96%

Concentration of the protein: 22 mg/ml

Crystallization



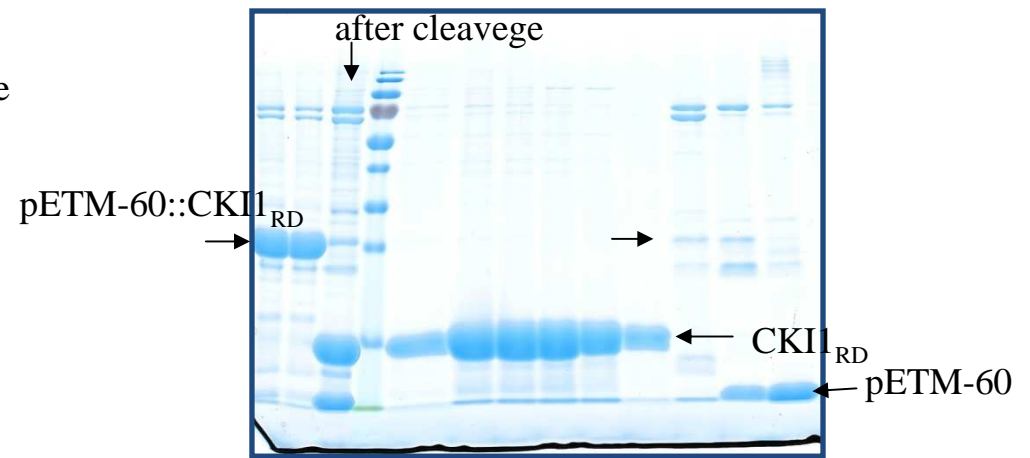
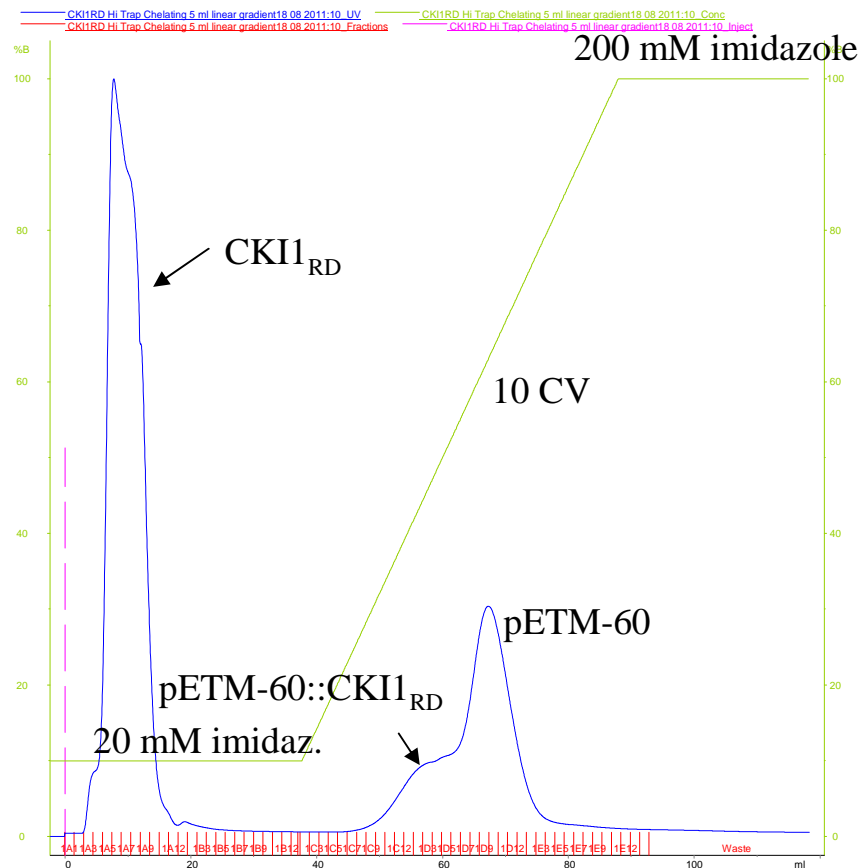
# His-tagged protein and IMAC under native conditions

## Five-step purification of *Arabidopsis* CKI1RD

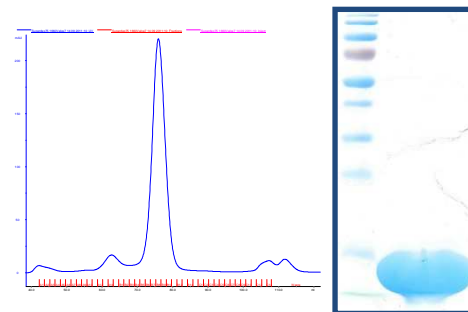
1. Affinity purification (MCAC)
2. Tag removal (TEV protease)
3. Affinity purification (MCAC)
4. Size exclusion chromatography



### 3. Affinity purification after TEV cleavage



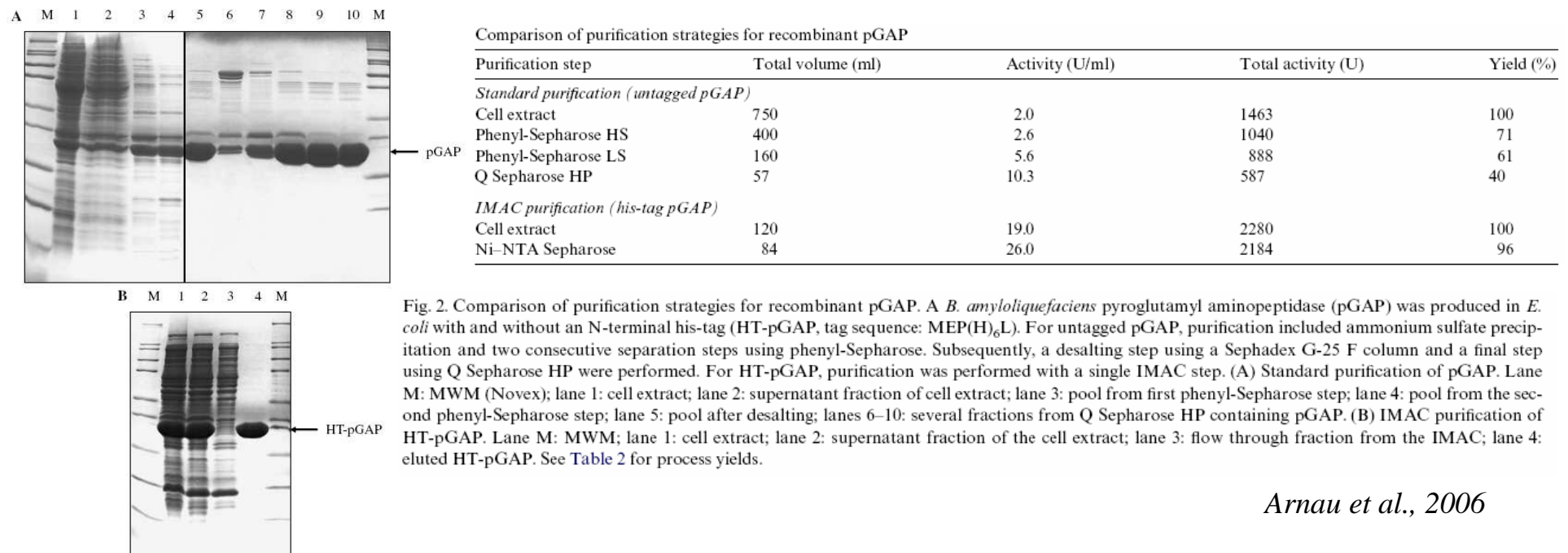
### 4. Size-exclusion chromatography





## Comparison of a standard purification process with affinity purification

- Generally, the yield and efficiency of any specific purification procedure depends on the level of optimization developed for individual proteins and the method. It is therefore recommended to use the data presented in different comparisons as indicative rather than definitive, which it is not e.g., identical elution conditions are optimal for different proteins.
- Standard chromatographic methods include several steps to obtain a relative pure protein. This results in a time-consuming procedure and a relatively low yield of recovery (typically 50 % of the starting material for optimized processes).
- The yields obtained in purification of proteins using affinity chromatography can be over 90 % and include a reduced number of steps.



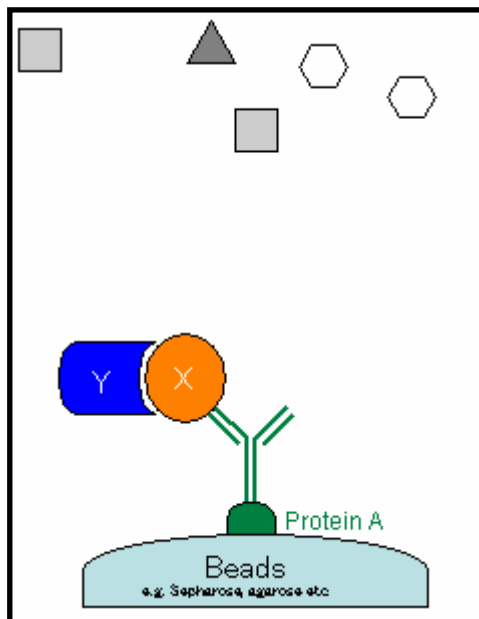


## Affinity purification for studying protein-protein interaction

- Affinity purification provides a high-efficiency method for isolation of interacting proteins and protein complexes:
  - **Co-immunoprecipitation**
  - **GST (or His) pull-down**
  - **Tandem affinity purification**
- Testing known protein-protein interaction.
- Identification of novel protein-protein interactions.

## Co-immunoprecipitation (Co-IP)

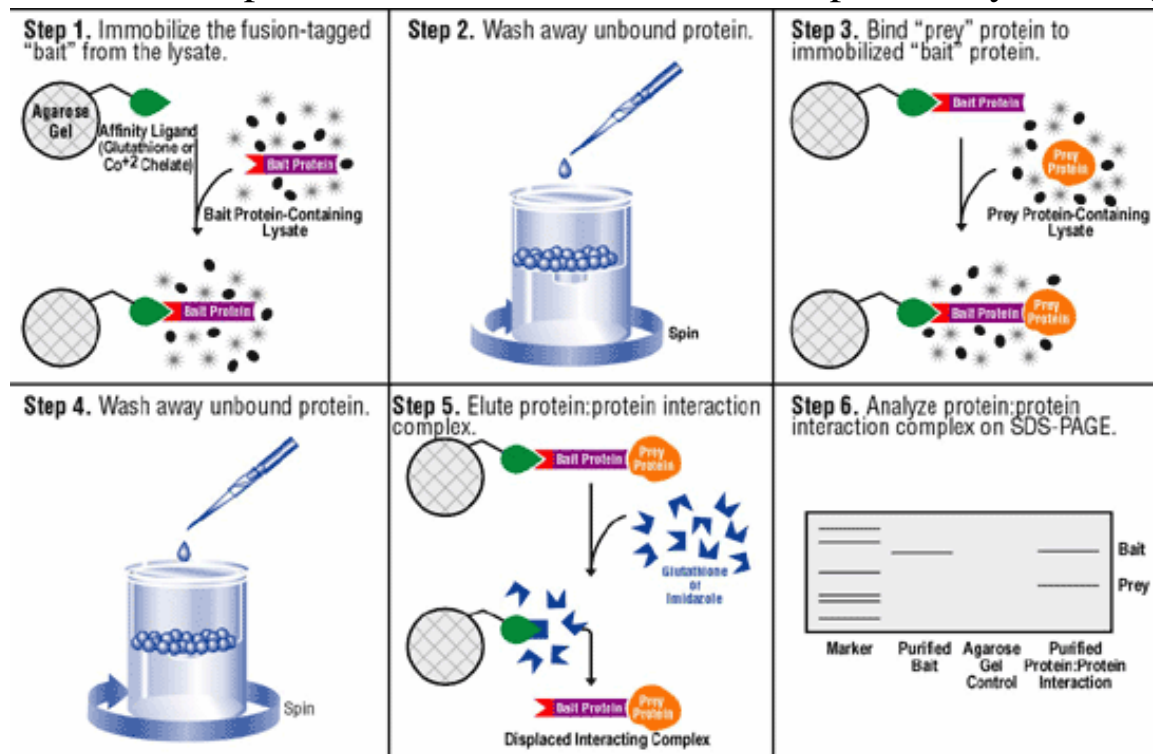
- The principle: If protein X is immunoprecipitated with an antibody of X, then protein Y, which is stably associated with X in vivo, may also be precipitated. **This precipitation of protein Y, based on a physical interaction with X, is referred to as co-immunoprecipitation.**
- An obvious advantage is that complexes are isolated in the state closest to the physiological condition.
- When a good quality antibody of X is available, Co-IP is a fast method and there is no need to clone and express the component(s) of the complex.



1. Cell lysis under mild conditions that do not disrupt protein-protein interactions (using low salt concentrations, non-ionic detergents, protease inhibitors, phosphatase inhibitors).
2. The protein of interest (X) is specifically immunoprecipitated from the cell extracts (using an antibody specific to the protein of interest or to its fusion tag).
3. The antibody-protein(s) complex is then pelleted usually using protein-A or G sepharose, which binds most antibodies .
4. Eluted immunoprecipitates are then fractionated by SDS-PAGE.
5. A protein of known identity is most commonly detected by performing a western blot or autoradiography when the interaction partner is labeled with  $S^{35}$  methionine. Identification of novel interaction is carried out by mass spectrometry analysis.

## Pull-down assay

- Pull-down assays are a common variation of immunoprecipitation and are used in the same way, but pull down does not involve using an antibody specific to the target protein being studied.
- They are used for purification of multiprotein complexes *in vitro*.
- The target protein is expressed in *E. coli* as GST fusion and immobilized on glutathione-sepharose beads (GST alone is often used as a control).
- Cellular lysate is applied to the beads or column, and the target protein competes with the endogenous protein for interacting proteins, forming complexes *in vitro*.
- Centrifugation is used to collect the GST fusion probe protein and adhering proteins.
- The complexes are washed to remove nonspecifically adhering proteins.



➤ Free glutathione is used to elute the complexes from the beads, or alternatively the beads with attached complexes are boiled directly in an SDS-PAGE sample buffer.

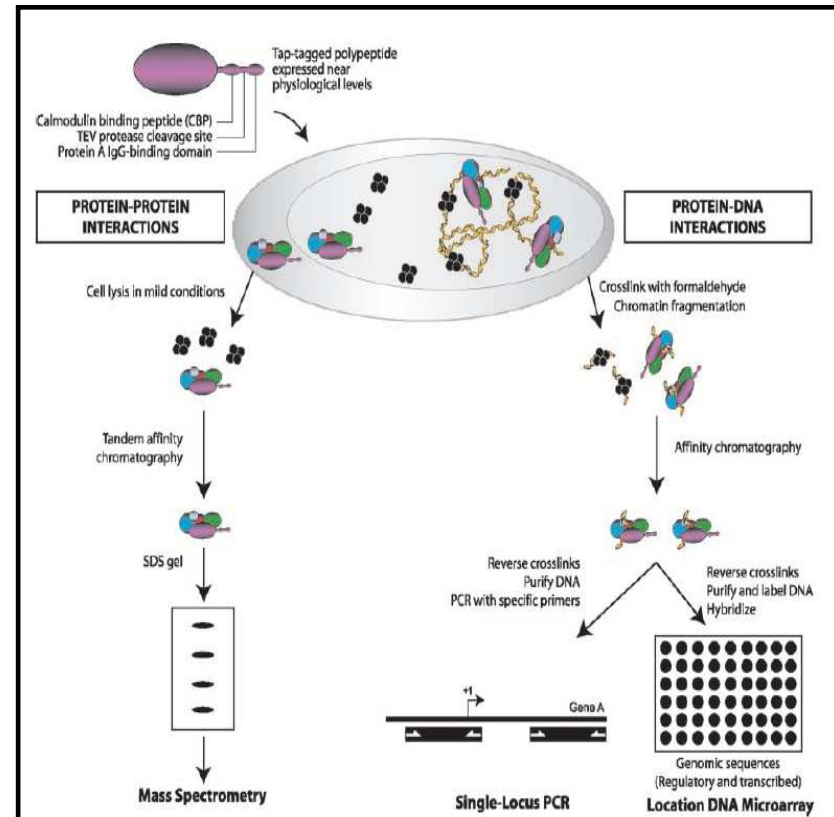
➤ The proteins are resolved on SDS-PAGE and processed for further analysis.

# Tandem affinity purification (TAP)

**Two-step purification strategy** in order to achieve higher purity of isolated multiprotein complexes under near physiological conditions.

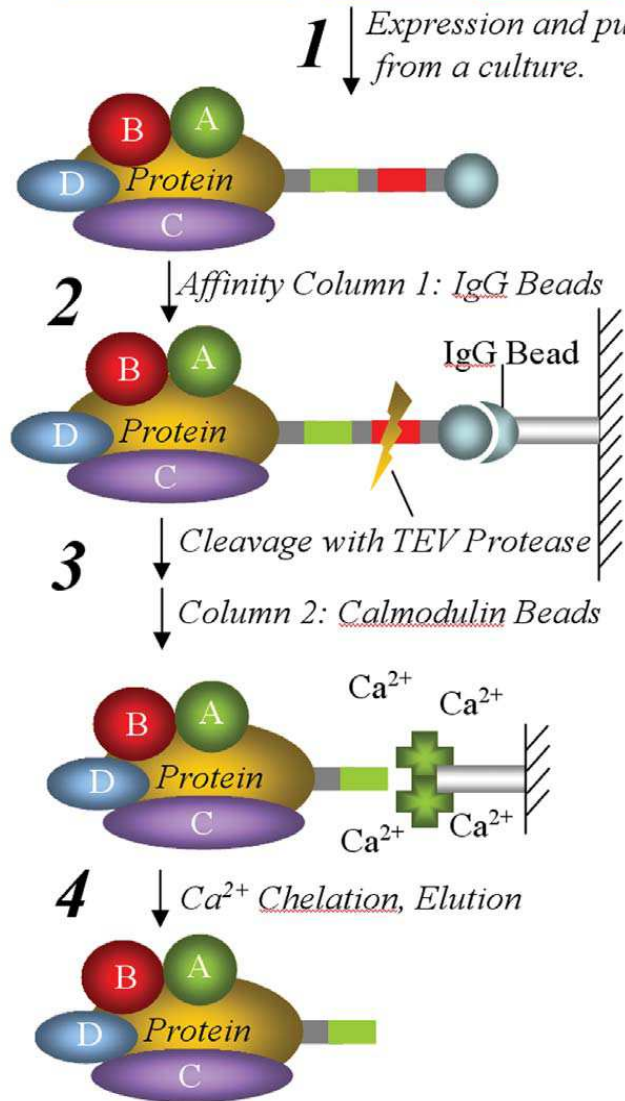
This method was originally developed for use in yeast and quickly adapted to higher eukaryotes such as insect cells, human cells and plant cells.

**Examples of TAP (tandem affinity peptides) tags**  
*TAP* tag: a double affinity tag (highly specific) which is fused to a protein of interest as an efficient tool for purification of native protein complexes.



- (a) **Protein A** - **Protein A** - **TEV** - **CBP** 20.7 kDa
- (b) **Protein G** - **Protein G** - **TEV** - **SBP** 18.8 kDa
- (c) **FLAG** - **StreptII** - **StreptII** 4.6 kDa
- (d) **RGS-6xHis** - **Biotin signal** - **6xHis** 9.6 kDa

# Tandem affinity purification



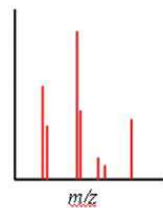
**5** Separation



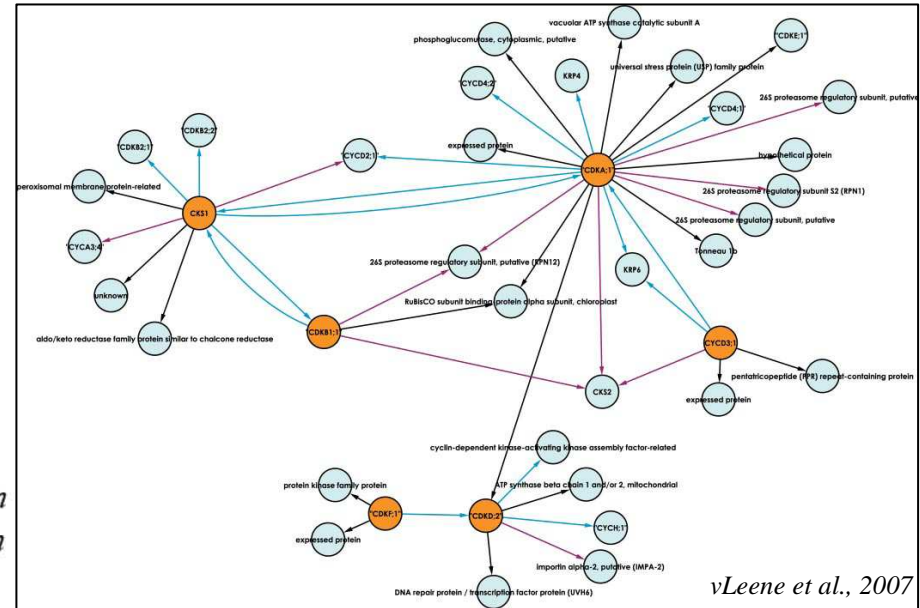
**6** Band Excision And Digestion



**7** Peptide Separation And MS Analysis



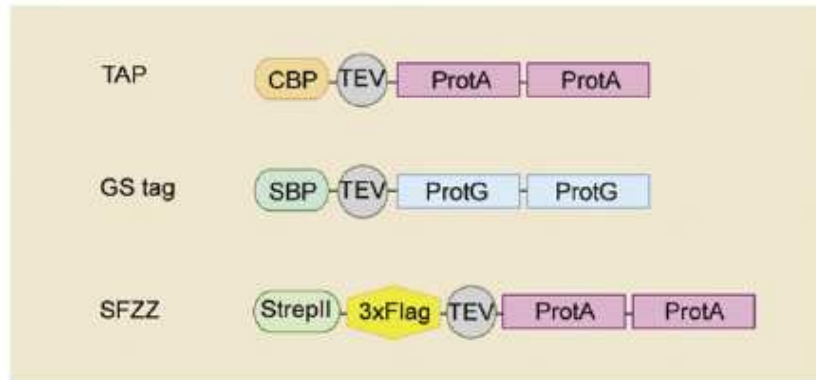
1. Protein
2. A
3. B
4. C
5. D



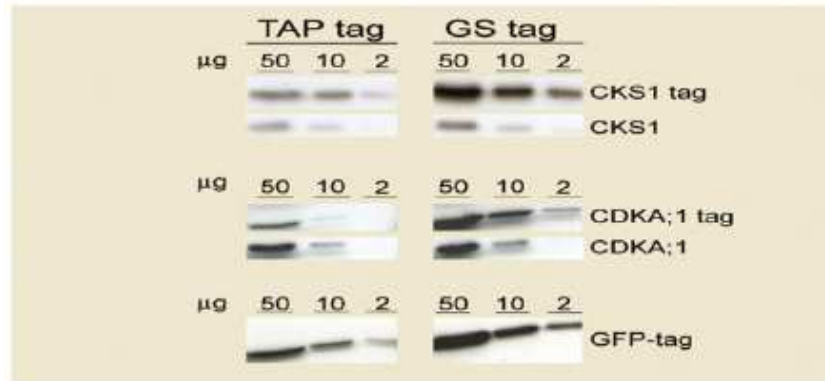


# A new (and so far the best) TAP tag for plants: The GS tag

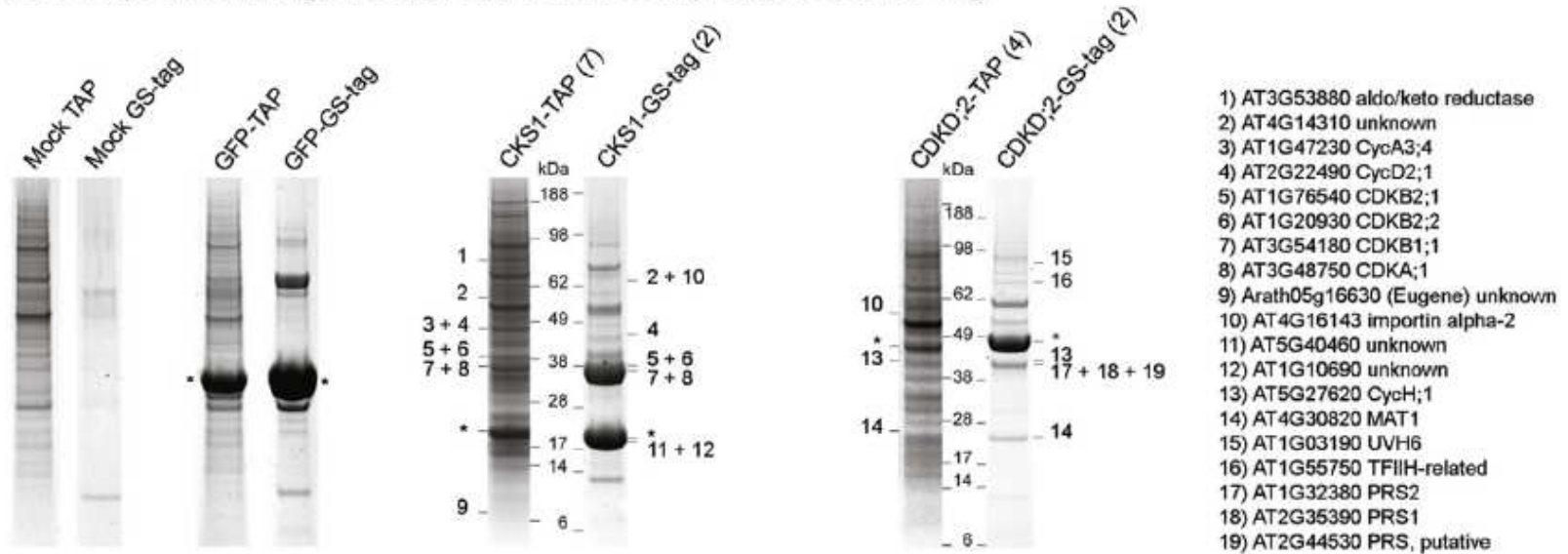
(a) Overview of tested TAP tags



(b) Higher bait expression with GS tag fusions



(c) Lower background and higher complex yield with GS tag compared to TAP tag



## **Affinity purification for studying protein-protein interaction**

- An affinity tags can influence protein-protein interactions (testing N- and C-terminal fusions).
- Loss of weak or transient protein-protein interactions.
- Non-specificity: controls, affinity tags with higher specificity
- Verification of newly identified interactors by other methods and biologically relevant mutants.

