Tags in protein expression, detection and purification

Lecture notes are available at:

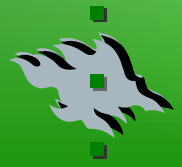
https://jeltsch.org/practical_molecular_biology

Registration for the practical course:

https://elomake.helsinki.fi/lomakkeet/52433/lomake.html

Complaints, suggestions, questions, etc.:

michael@jeltsch.org



What are protein tags?

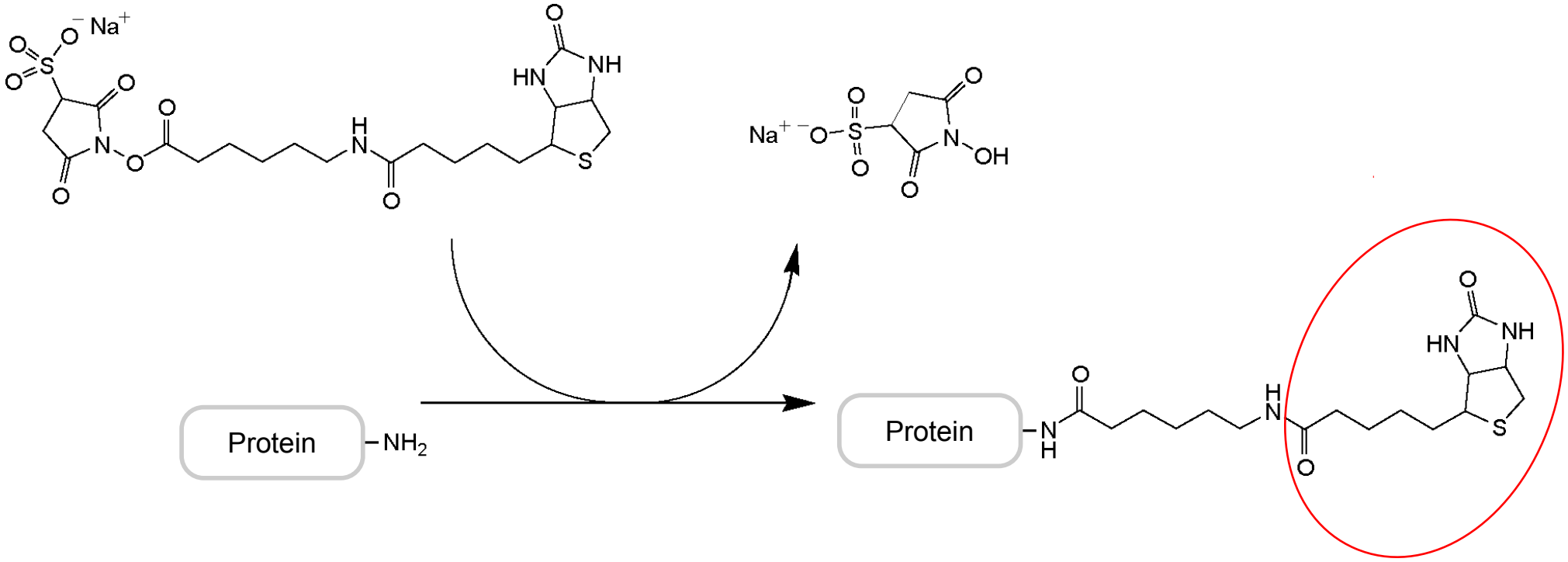
- Rather “short” polypeptide sequences which are attached to a protein to confer a specific property.
- Attached via modification of the DNA sequence (as opposed to labels, which are attached post-translationally via chemical modification).
- This definition is not used uniformly!

Examples: His-tagging vs. biotinylation

Hexahistidine tag:

```
...GCA CAT CAT CAC CAT CAC CAC TGA
... A   H   H   H   H   H   H   *
```

Biotinylation:



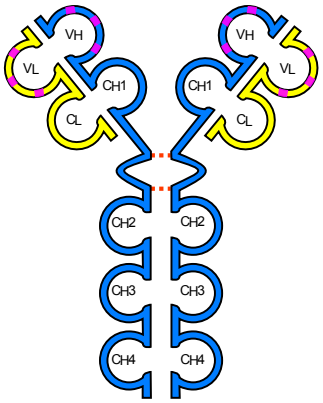
"Chemical" biotinylation requires a fairly pure protein!

Why would you tag?

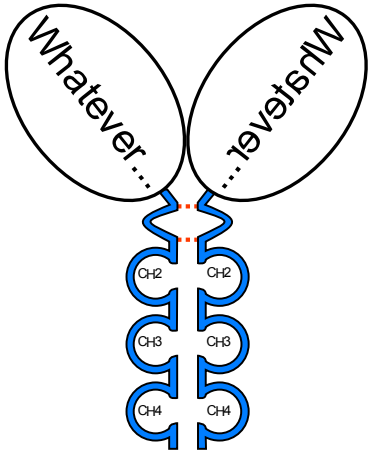
Because you want to modify the physical and chemical properties of your protein

- If you don't have a specific antibody (detection tag)
- If you want to purify it (purification tag: affinity purification, ion exchange)
- If you want to label it without the need to purify it (labeling tag)
- If you need to increase the solubility
- If you want to change the stoichiometry (e.g. dimerization)

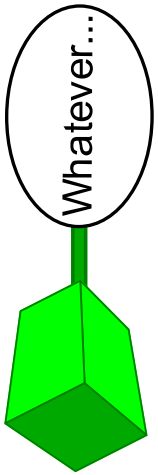
What is the difference between a tag and a fusion protein?



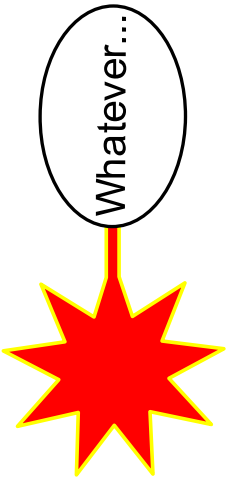
IgG_1



F_C



GFP



GST

kDa		26	27	26

$-H_6$	~0.84 kDa
Biotin	~0.24 kDa

~ Tags too small to be stably expressed on their own. Small peptide expression requires a fusion partner (possibly in tadem with met spacers for CNBr cleavage)!

Size matters! Bigger is not always better.

Protein Pharmaceuticals

- Jenner, 1796: First “protein vaccine” cow-pox
- Banting & Best, 1922: First protein pharmaceutical insulin
- Genentech, 1978: human insulin in E. coli (FDA approval in 1982)
- Genentech, 1987: FDA approval of first recombinant vaccine (Hepatitis B vaccine) replacing the human blood-derived vaccine
- Today more than 200 approved peptide and protein pharmaceuticals on the FDA/EMA lists

What properties of proteins can be exploited for purification?

- **Size (“hydrodynamic radius”)**
  → Gel filtration (GF, SEC)
- **Solubility**
  → Differential precipitation
- **Surface charge (hydrophobicity)**
  → Hydrophobic Interaction Chromatography (HIC),
  → Reverse Phase Chromatography (RPC)
- **Affinities**
  → IMAC – Histag
  → Protein A – antibody
  → carbohydrates – Concanavalin A
  → Protein ligand – cognate receptor
  → Specific antibody – cognate antigen

Protein Pharmaceuticals

- List of all EMA-approved medicines (purification “details” in the scientific discussion
  http://www.ema.europa.eu/htms/human/epar/a.htm
- List of all FDA-approved medicines
  http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.c
- (Detailed) description of purification is often not available (“trade secrets removed fr
  publicly available documents”)

No tags for protein pharmaceuticals

- The proteins need to be authentic (untagged) in the final product
- The FDA doesn't like tags: tags serve no purpose in the final product
- Potential of modifying the biological activity

  Immune response to a self-peptide triggered by tagging overriding tolerance

  → Tag needs to be removed

  → Conventional column chromatographic purification (ion exchange, gel filtration)

Affinity tag		*ligand*	*typical binding capacity/ ml*	*elution*	*Length (aa)*	*sequence*	*place-ment*
Histidine (His)	X	IDA-coordinated Me2+ (iminodiacetic acid)	10-40 mg	imidazol, low pH	5-15 (6)	HHHHHH	N, C, int
		NTA-coordinated Me2+ (nitrilotriacetic acid)		imidazol, low pH			
		CM-Asp-coordinated Me2+ (carboxymethylated aspartic acid)		imidazol, low pH			
Strep II	X	Strep-Tactin	3 mg	d-desthiobiotin	8	WSHPQFEK	
Strep III	X	Strep-Tactin	1.5 mg	d-desthiobiotin	28	WSHPQFEKGGGSGGGSGGGSWSHPQFEK	
Streptavidin binding protein (SBP)		streptavidin		biotin	38	MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP	
T7		mAb	0.3 mg	low pH	11/16 (11)	MASMTGGQQMG	N, int
FLAG	X	mAb	0.6 mg	low pH, FLAG peptide	8	DYKDDDDK	N, C, int
Ribonuclease S peptide (S)		Ribonuclease S protein, mAb		low pH	15	KETAAAKFERQHMDS	N, C
Softag1		Polyol-responsive mAbs		polyol	13	SLAELLNAGLGGS	
Softag2		Polyol-responsive mAbs		polyol	8	TKDPSRVG	
Elastin		none		temperature shift	18-320		
Haemagglutinine (HA)		mAb	2-3 mg	low pH	9	YPYDVPDYA	N, C
C-Myc	X	mAb		low pH	11	EQKLISEEDLN	N, C, int
V5		mAb		low pH	14	GKPIPNPLLGLDST	
Xpress		mAb		low pH	8	DLYDDDDK	N
Dihydrofolate reductase (DHFR)	X	Immobilized methotrexate (MTX)		MTX	171		N
Chitin binding domain (CBD)		Chitin	2 mg	DTT	51		N, C
Calmodulin binding domain (CBD)	X	Calmodulin	1 mg	EGTA, EDTA	26	KRRWKKNFIAVSAANRFKKISSSGAL	N, C
Immunoglobulin G Fc (IgGFc)	X	Protein A	50 mg	low pH, high salt, ethylene glycol	232		C
Protein A	X	Immunoglobulin G	2 mg	low pH, high salt, ethylene glycol	122		N
Cellulose binding domain (CBD)	X	Cellulose		H2O, high pH, urea, ethylene glycol, guanidinium HCl	27-129 (108)		N, C
Glutathion S-transferase (GST)	X	Glutathione, mAb	10 mg	glutathione	201-220		N
Maltose binding protein (MBP)		Amylose	6-8 mg	maltose	396		N, C
Octaarginine (Arg8)		Ion exchange resins		high salt	8	RRRRRRRR	N, C
AviTag		BirA (bacterial biotin ligase)			15	GLNDIFEAQKIEWHE	N, C, int

Affinity tag		*kDa*	*Comments*	*Major supplier of medium*
Histidine (His)	X	0.84		GE Healthcare
				Qiagen
				Clontech/BD Biosciences
Strep II	X	1.06	modified streptavidin	IBA, Qiagen
Strep III	X	2.87	tandem strep II tag	IBA, Qiagen
Streptavidin binding protein (SBP)		4.3		Thermo Scientific/Pierce
T7		1.1	amino terminal 11 aa of the T7 gene 10 protein	Merck, Thermo Scientific/Pierce, Abcam
FLAG	X	1.01	Ca2+ dependend; mAbs recognize different placements of the tag (e.g. the M1 mAb recognizes only N-terminal FLAG)	Sigma
Ribonuclease S peptide (S)		1.75	derived from bovine pankreatic ribonuclease A by subtilisin cleavage	Bethyl Laboratories
Softag1		1.2	recognized by polyol-responsive mAbs	Neoclone, Lucigen?
Softag2		0.86	recognized by polyol-responsive mAbs	Neoclone, Lucigen?
Elastin			protein aggregation by temperature shift	
Haemagglutinine (HA)		1.1	influenza virus derived, more suitable for detection than for purification	Sigma, Roche, Babco
C-Myc	X	1.32		Sigma, Babco
V5		1.42	more suitable for detection than for purification	Abcam
Xpress		1	more suitable for detection than for purification	Invitrogen
Dihydrofolate reductase (DHFR)	X	19		
Chitin binding domain (CBD)		6	with the intein (self-cleavable tag) portion the tag size is 28 kDa (C-terminus) or 56 kDa (N-terminus)	NEB
Calmodulin binding domain (CBD)	X	2.96	binding is Ca2+ dependend	GE Healthcare
Immunoglobulin G Fc (IgGFc)	X	26	increases biological half-life, dimerizes the fusion partner	GE Healthcare, Thermo Scientific/Pierce
Protein A	X	14		GE Healthcare
Cellulose binding domain (CBD)	X	11	suitable for immobilization	Novagen
Glutathion S-transferase (GST)	X	26	enhances solubility	GE Healthcare, Thermo Scientific/Pierce
Maltose binding protein (MBP)		40	enhances solubility	NEB
Octaarginine (Arg8)		1.27		GE Healthcare, Toso
AviTag		1.83	For specifically labeling Avi-tagged proteins intracellularly	Avidity

Histag is the only possible tag for large scale purification of proteins

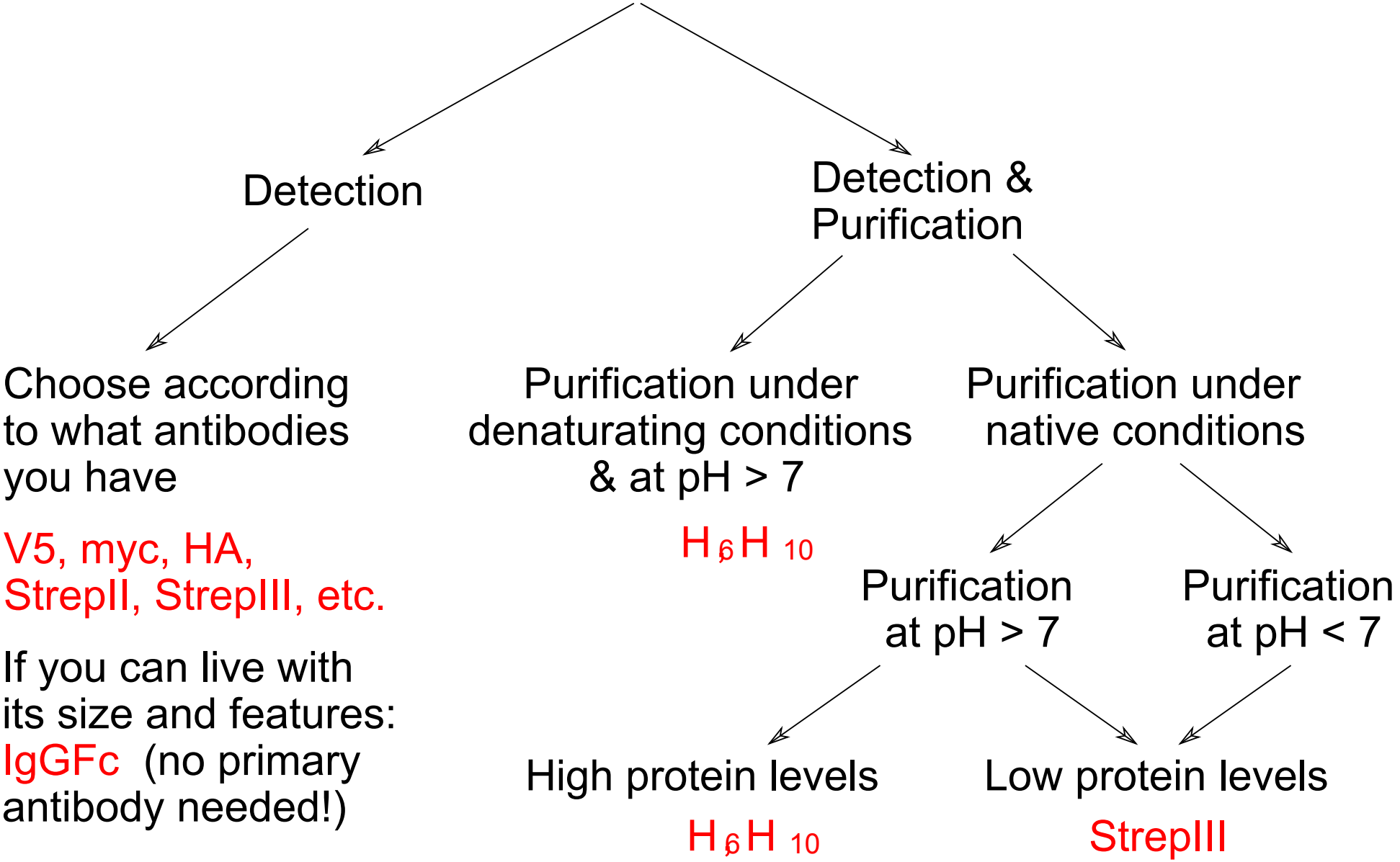
- Small tag size, “relatively” non-immunogenic
- Stability of the affinity resin towards high pH & detergents
- One of the lowest price/binding capacity ratio
  (to purify 1g VEGF-C, the affinity matrix costs are ~200€ for Ni^{2+}NTA Superflow and ~5000€ for streptactin)
- No patent restrictions anymore since 2005

Still too expensive

- Cost/binding capacity ratios: IEX 4-5 cent/mg vs IMAC 18 cent/mg)
- 270$ vs. 1080$ media costs for a single typical Rituximab/MabThera* regimen (1x0.75g + 5x1g
- Production in CHO cells, purification by Protein affinity and AIEX chromatography

*Best-selling mAb of the last decade ~ 2 billion $/year, anti-CD20 mAb for the treatment of non-Hodgkin's lymphoma, chronic lymphocytic leukaemia and rheumatoid arthritis

What tag should I use?



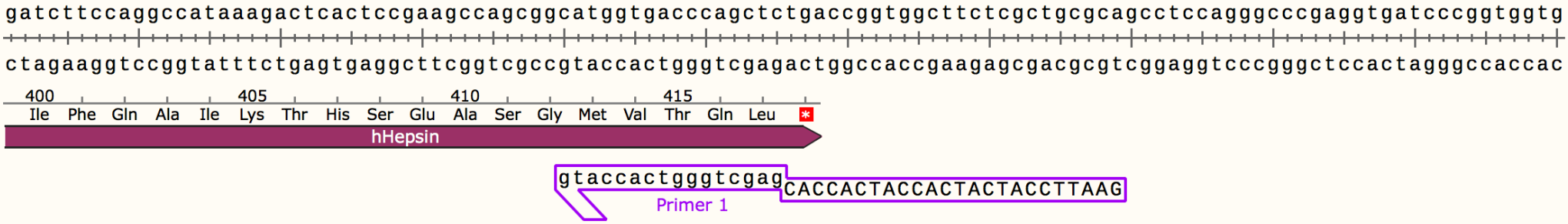
What should I use?

- Purification: His_6, His_{10}, StrepII, StrepIII, IgGFc*
- TAP tags (Tandem Affinity Purification, e.g. StrepII-TEV**-His_6 or CBP***-TEV-ProteinA), originally invented for interaction analysis
- Cleavage sites (TEV protease) or alternative location (C-terminus versus N-terminus) if the tag causes problems
- Tags for detection: Same as for purification and many more (V5, myc, Flag...)
- Same tag, different name: StrepIII = twin strep, FLAG = DKK
- Watch out: V5 (Simian Virus 5), myc (aa 408-439 of hMyc), HA (flu virus), Xpress tag (bacteriophage T7 gene 10 protein), ...

*IgGFc has biological activity (it binds Fc receptors on immune cells), it dimerizes the protein and increases its biological half life, **Tobacco Etch Virus ***Calmodulin Binding Protein

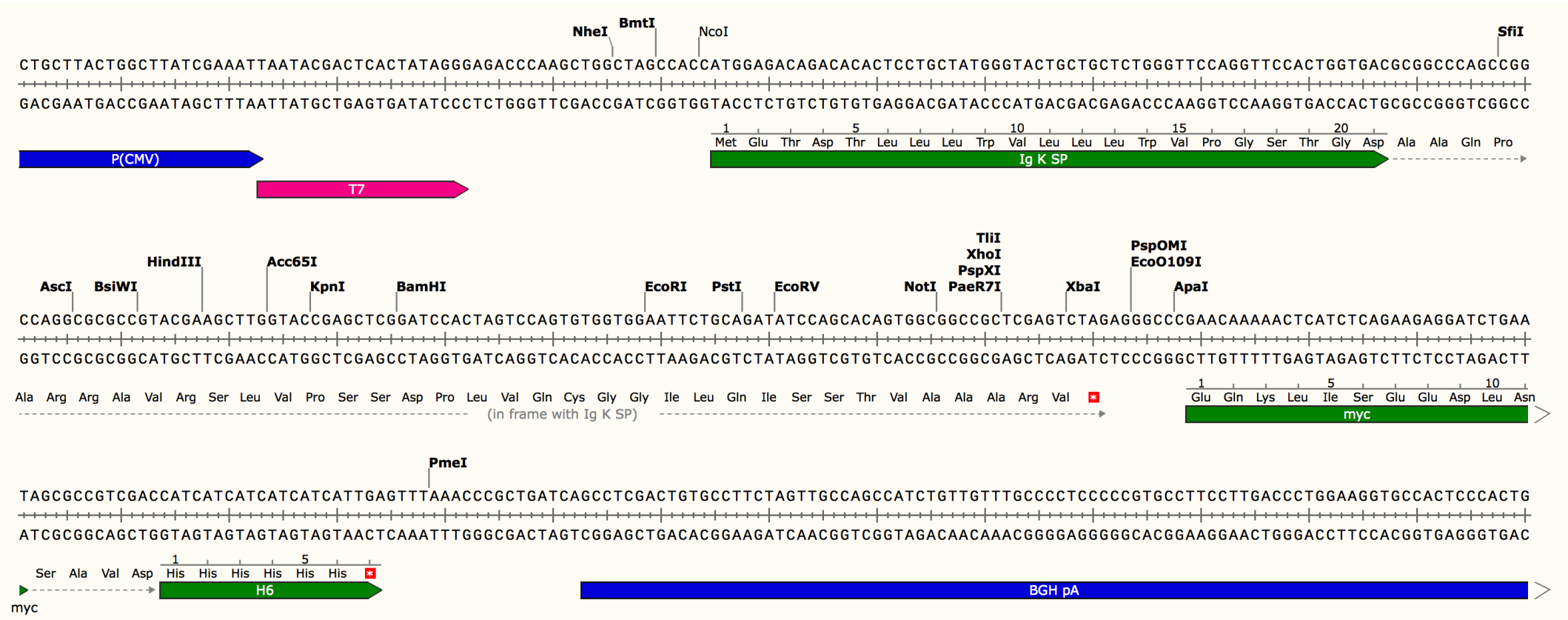
How to attach a tag to a protein?

- Short tags: By PCR with a tailed PCR primer

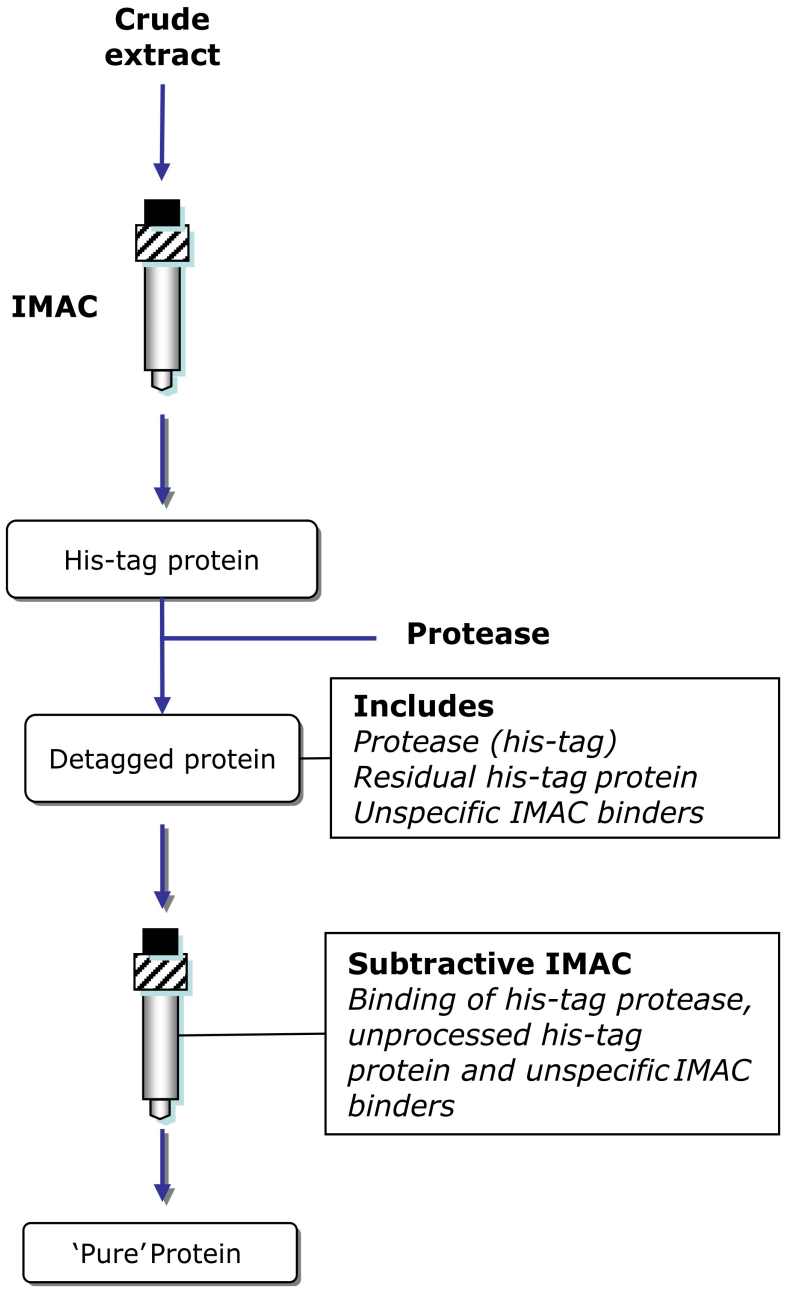


How to attach a tag to a protein?

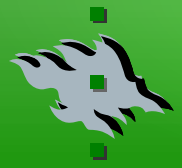
- Long tags: Dedicated vector that contains the tag-sequence (many commercial vectors contain common tag sequences after of before the MCS)



Requires enzymatic cleavage of protein and tag removal



- Additional affinity column
- Proteases are expensive and not very stable
- Impurities (Ni^{2+}, un-cleaved and hetero-genously cleaved protein, protease)



What is available from us

- http://research.med.helsinki.fi/corefacilities/akt/index.html
- http://akta.jeltsch.org