From cloning to protein analysis

Methods and strategies

Day 3

Expression strategies

Andrea Balan (abalan@usp.br)
Marko Hyvonen (mh256@cam.ac.uk)

Universidade de São Paulo
Brasil

UNIVERSITY OF CAMBRIDGE

Fusion constructs

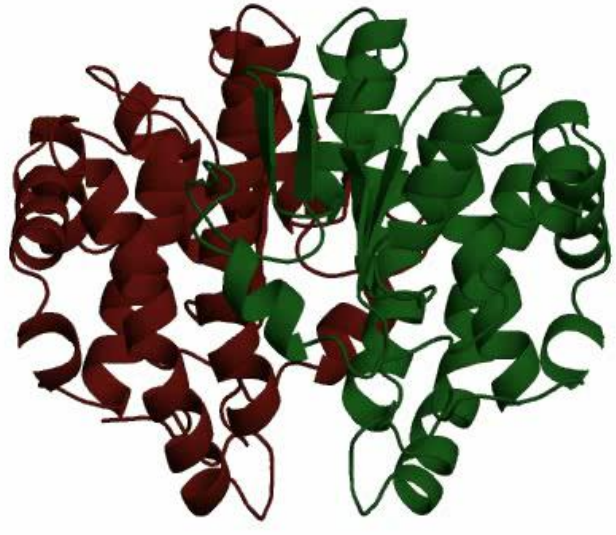
Common fusion partners

Fusion proteins:

- Glutathione S-transferase (GST) from *S. japonicum*
- Maltose binding protein (MBP) from *E. coli*
- Thioredoxin from (Trx) *E. coli*
- GB1 domain from *Streptococcus*
- β-lactamase from *Chromohalobacter* sp. 560
- SUMO, NusA, Streptavidin,….

Glutathione-S-transferase (GST)

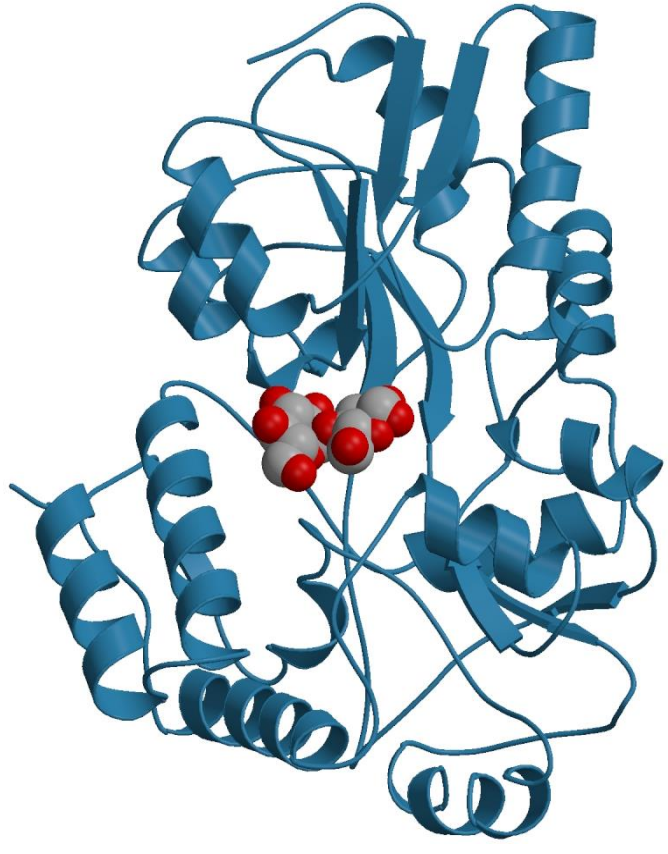
- pGEX vectors from GE Healthcare
  - *tac* promoter, pBR322 based
- ca. 30 kDa
- good purification tag
- dimeric, which can cause problems
- cysteines prone to cause disulfide linked aggregation
  - keep properly reduced !
    - Fresh (!) DTT in all buffers



PDB:1UA5

Maltose binding protein (MBP)

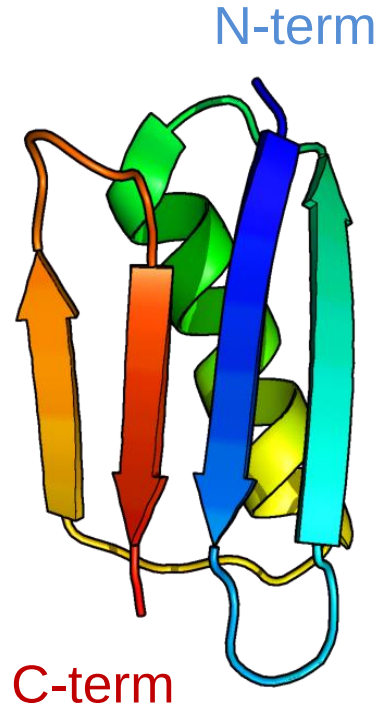
- pMAL vectors by NEB
  - *tac* promoter, pBR322 background
- monomeric protein of 43 kDa (with a His-tag)
- both intracellular *and* secreted versions (in pMAL
- relatively good at promoting solubility
- works also as an affinity tag
- Variants for crystallisation enhancement and with increased maltose affinity



PDB:1MDQ

GB1 domain

- Small domain from protein G, an IgG binding protein from group G *Streptococci*
- Only 7.5 kDa, ~9.5 kDa with a His-tag
- Fast folding protein
- Highly expressed
- Very soluble
- Good as a fusion for making peptides and small proteins

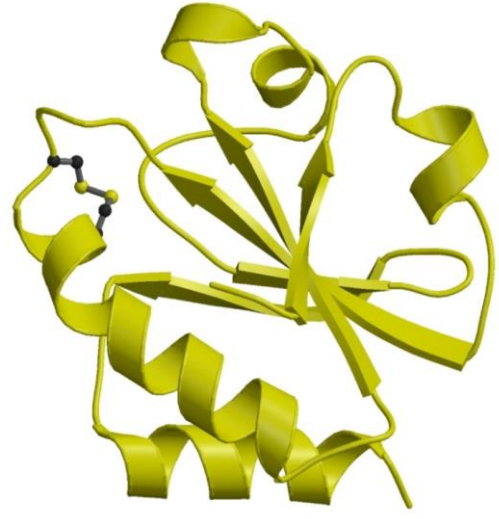


pPEPT1 plasmid

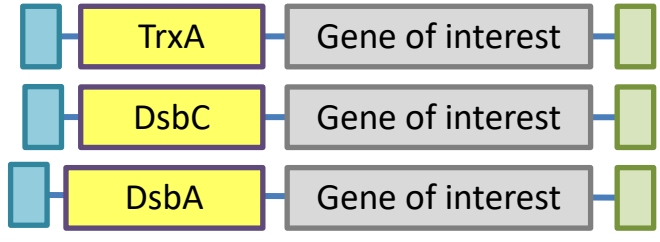
- Vector for making peptides in *E. coli*
- N-terminal StrepTag-GB1 fusion
- C-terminal His_8-tag
- TEV site for proteolytic cleavage
- "DP" sequence before and after the peptide for acid cleavage

Thioredoxin (TrxA)

- Disulfide isomerase from *E.coli*
- ~100 residues only, so one of the smaller fusion proteins *(14kDa)*
- Can be useful when expressing disulfide linked proteins in the cytoplasm of *E.coli*
  - *In combination with trxB/gor mutants: Origami(DE3) / OrigamiB(DE3) / SHuffleT7*
- Not for affinity purification
- Worth testing in parallel with DsbA and DsbC fusions

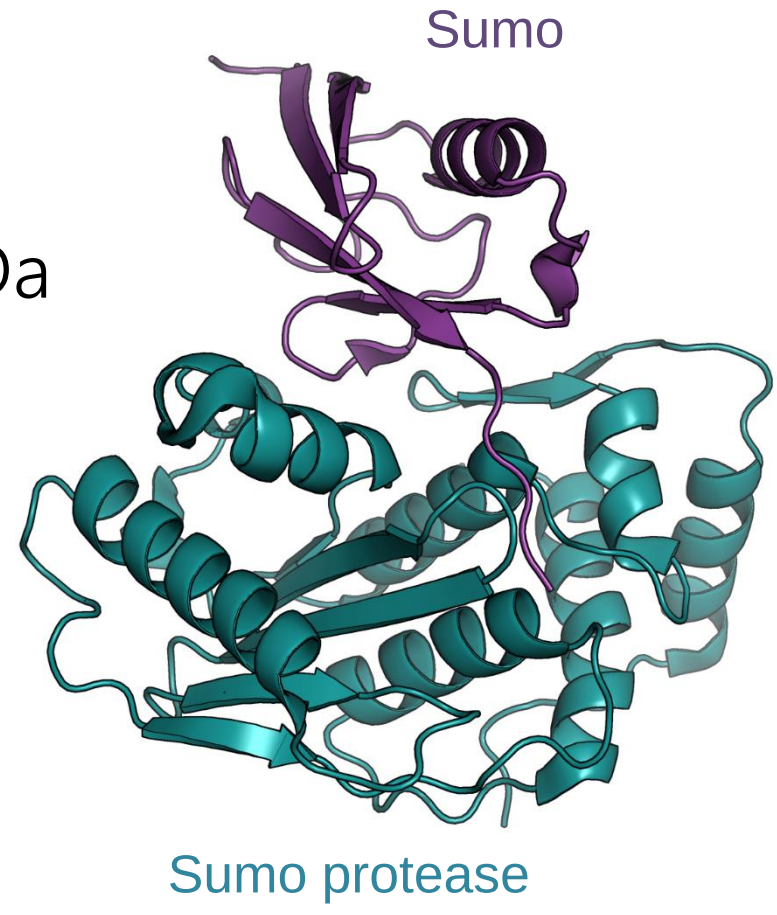


PDB:2TRX



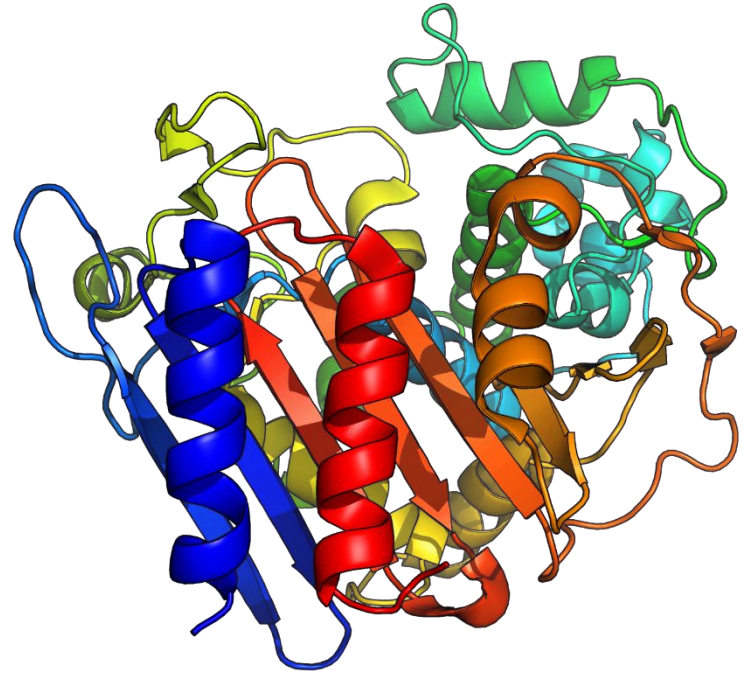
Sumo

- Small Ubiquitin-related modifier
- Small fusion, just 11 kDa (12 kDa with a His-tag)
- Folded Sumo recognised by the protease, ULP-1
- Cleavage in specific position with respect to Sumo fusion
- A very soluble protein



Halophilic β-lactamase (Bla)

- New kid on the block
- Acidic (pI 4.3) 42 kDa protein
- Seems to promote solubility very well
- No affinity matrices available
- Large fusion (like MBP)
- In our pExp fusion vectors



Bla

Linear peptide tags

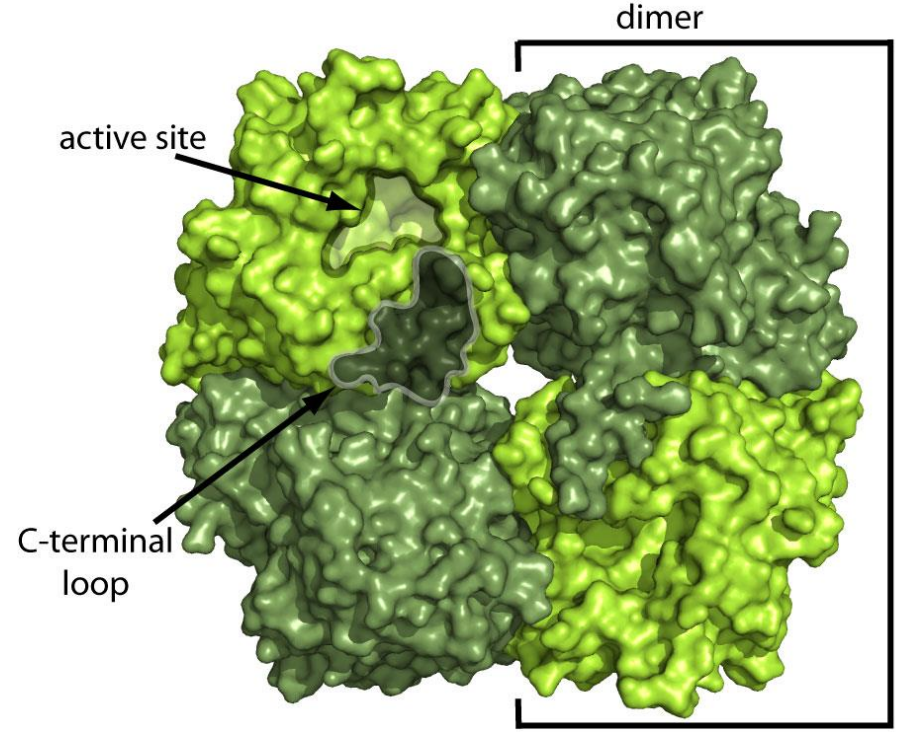
- Relatively short sequences
  - Usually not affecting the properties of the target
  - Are natively unstructured, and can have negative effect on crystallisation
  - Optimised codons give good translational start and can boost expression
  - Easy to introduce by PCR even if plasmid would not have one
- **Common fusion peptides/tags**
  - His-tag, $(His)_6$, hexa-His, (octa-His, deca-His)
  - StrepTagII
  - Avi-tag
  - S-tag, FLAG-tag, myc-tag, …

His-tag

- Developed at Hoffmann-La-Roche in the late '80s
- Based on affinity of adjacent histidines on immobilised metal ions
- Independent on position, as long as accessible
  - N-terminal, C-terminal, in the middle
- Typically 6 histidines used, but even two are already useful
  - Some proteins bind strongly to NiSepharose without a His-tag
- Removal not always necessary, as seldom affects the function of the protein

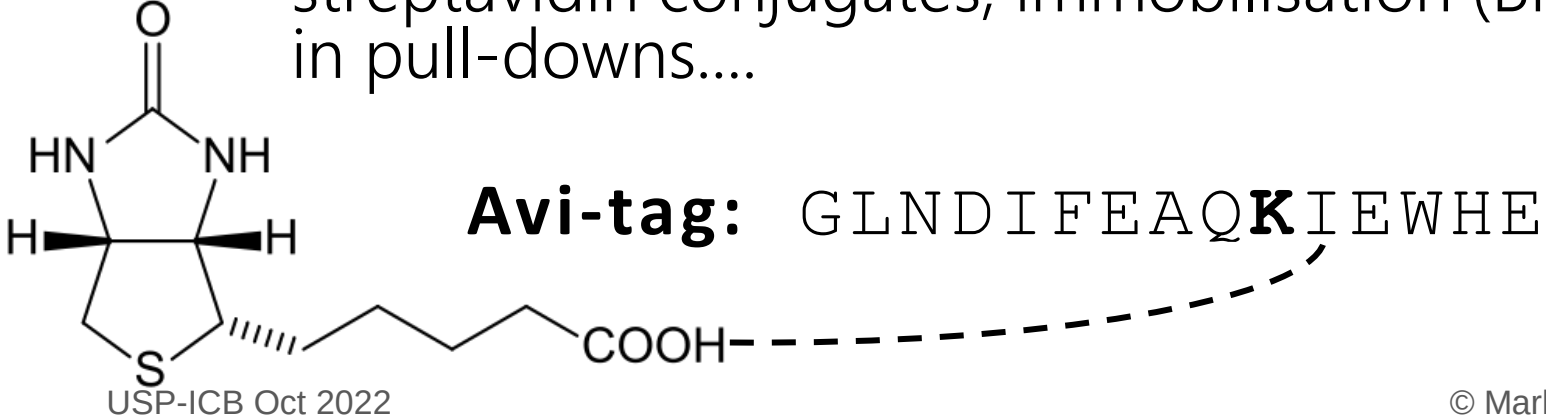
Strep-tag II

- Developed and sold by IBA
- Octa-peptide “W S H P Q F E K”
- Can be fused N- or C-terminally
  - Easily incorporated using PCR
- Binds modified streptavidin, Streptactin
- Highly specific purification in one step
- Elution with desthio-biotin
  - Very mild conditions
- Column material on the expensive side



Avi-tag

- 15 amino acid tag that can be placed in the N- or C-terminus
- Recognised by *E. coli* biotin-transferase enzyme BirA which adds a single biotin to it with 85-95% efficiency
  - Defined site of biotinylation, unlike with chemical biotinylation which can be less accurate.
- Biotinylation can be done in vivo, by co-expressing BirA, or in vitro using purified BirA enzyme.
- Biotinylated protein will bind to streptavidin like any other biotin containing protein: detection by streptavidin conjugates, immobilisation (Biacore), use in pull-downs….



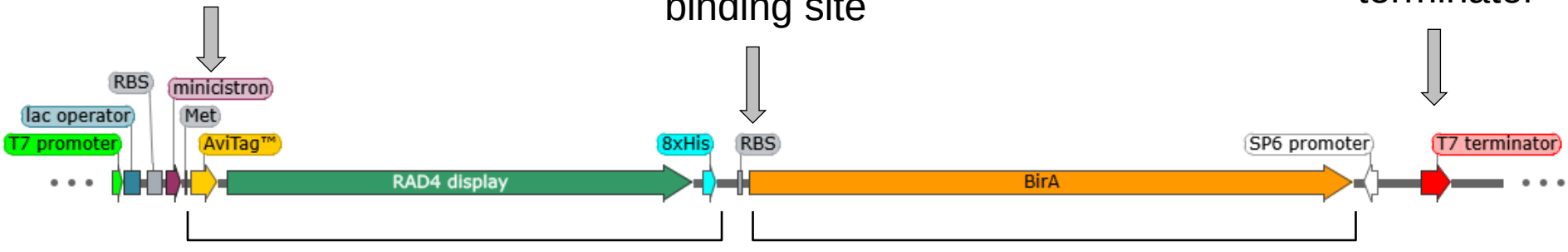
Avi-tag: GLNDIFEAQ**K**IEWHE

pRAD v.2

Mini-cistron for better expression

Second ribosome binding site

Single transcription terminator



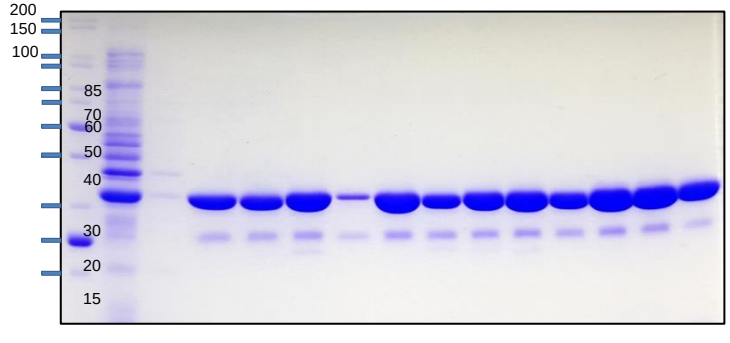
RAD display protein with:

- N-terminal Avi-Tag for labelling
- C-terminal His_8-tag
- SLIC cloning site

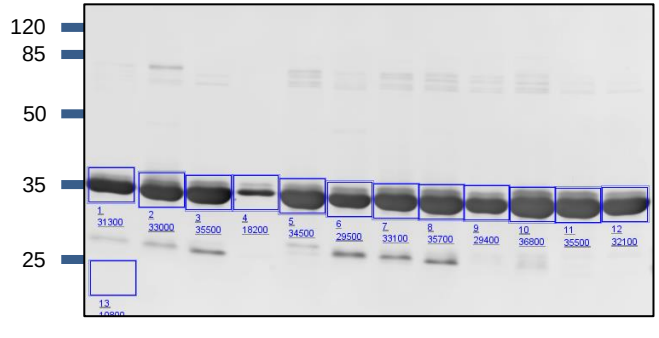
Co-expressed BirA

- Biotin transferase that modifies the Avi-Tag
- Labelling of the target during expression

Epitope library purified by Ni-NTA



Western blot with streptavidin probe

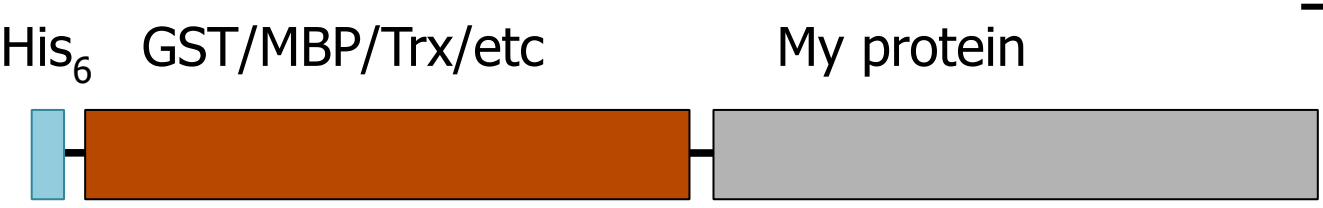


Effect of fusion on the final protein yield

	Target MW	Fusion MW	MW Ratio	Yield after tag removal
MBP	20 kDa	40 kDa	1:2	33 %
GST	20 kDa	30 kDa	2:3	40 %
GB1	20 kDa	10 kDa	2:1	66 %
His_6	20 kDa	2 kDa	10:1	91%
No fusion	20 kDa	0 kDa	n/a	100 %

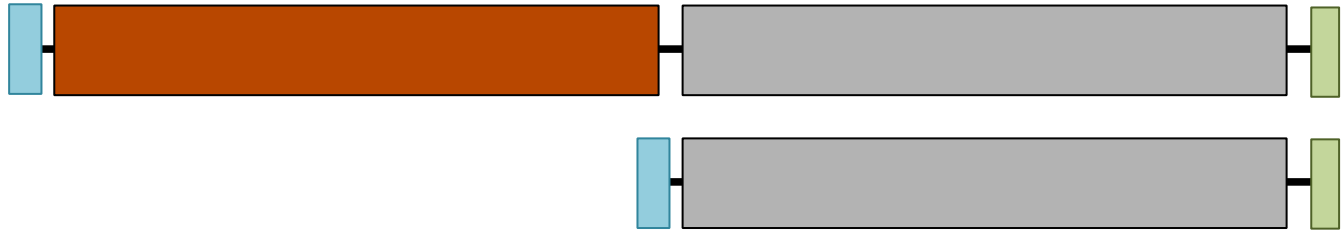
Assuming equal expression level (mgs) for every construct

Double affinity tags



His-GST: double affinity purification
His-MBP: double affinity purification
His-Trx: Trx to solubilise the protein, His-tag for purifications

All pBAT, pOP pExp vectors



All pExp vectors

N-terminal and C-terminal double-tagged protein to pull down full-length proteins

Are fusions only good for you ?

- Fusions can offer several advantages for expression work, but have also some disadvantages
  - Purification depends only on the fusion partner, not your protein
  - Yield of your protein can be very small after proteolytic release and further separation
  - Solubility of the fusion protein might mislead you into thinking that all is well with your protein
- Do not forget conventional purification techniques and expressing proteins without fusions.

Why express in *E.coli* ?

- A very well established system
- Easy to manipulate
- Fast growing – lab-scale experiments done in one day
- Large variety of vectors, strains, methods
- Low-tech, safe and inexpensive to grow
- Allow for scale up to industrial level expression
- Suitable for variety of labellings
  - isotopes for NMR (^{13}C, ^{15}N, ^{2}H)
  - non-natural amino acids (Se-Met & Se-Cys for crystallography)
  - radioactive (^{35}S, ^{14}C, ^{3}H)

And why not ?

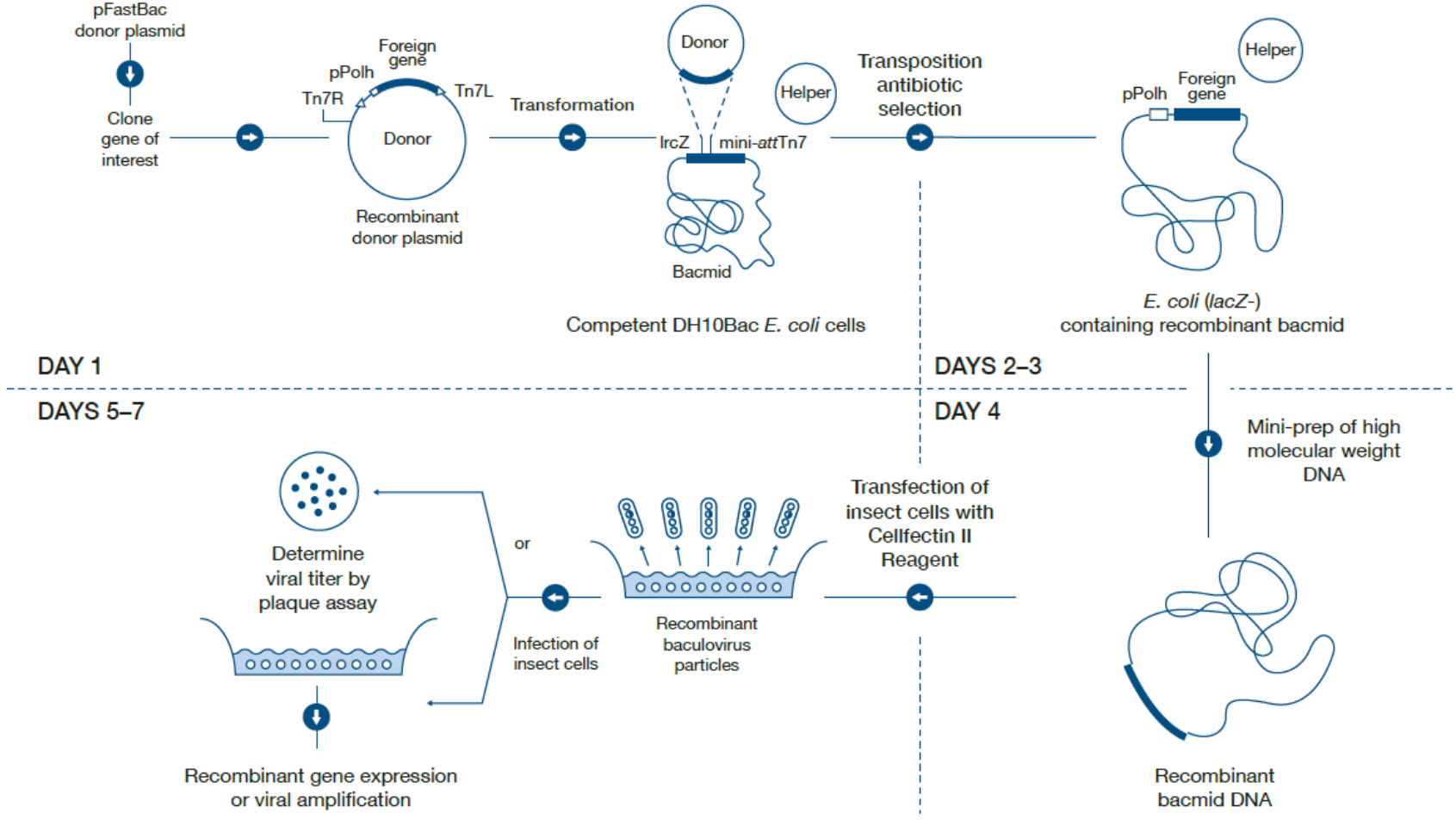
- Eukaryotic post-translational modifications are missing
  - glycosylation, disulfide bridges, lipidation, proteolytic processing etc.
  - (could be a blessing too…)
- Often problems with solubility of proteins
  - Natural chaperones are missing
  - Too high level and speed of expression compared to the rate of folding of the protein of interest
  - Proteins expressed in isolation, without their natural binding partners

What else could you use?

- Insect cells using baculoviruses
- Mammalian expression: HEK293, CHO cells
- Yeast expression: *, P. pastoris, S. cerevisiae*
- In vitro transcription translation:
  - Wheat germ extract
  - *E. coli* extract
- Others: filamentous fungi, algae, plants…

Insect cells/baculovirus

Expression in Lepidoptera (moth/butterfly larva) cells using viral delivery
Cells grow at 28°C and not need CO_2, unlike mammalian cells

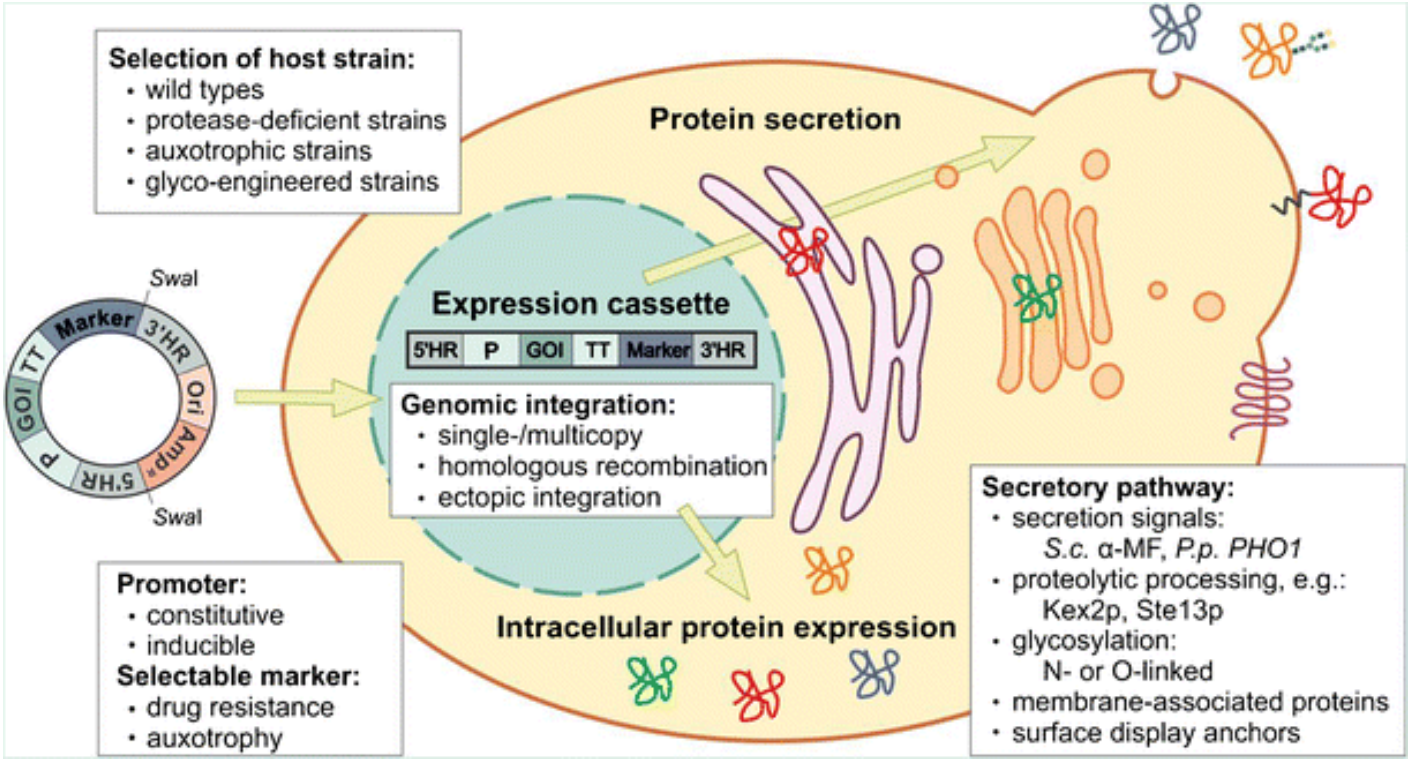


Mammalian cell expression

- Most often CHO and HEK293 cells
- Transient or stable expression
  - testing usually as transient expression (few days)
  - selection of clonal cell line under antibiotic selection (weeks)
- CHO cells typically used for large scale antibody production of antibodies
- Very good for secretion of bioactive proteins
  - all the mammalian post-translational modifications
- Requires dedicated facilities
  - 37°C incubation, CO_2 atmosphere
- Slower growing cells

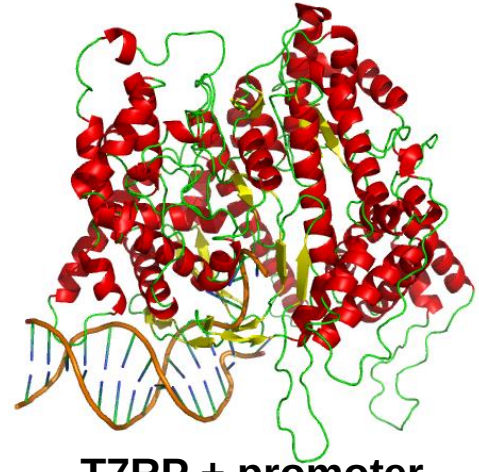
Pichia pastoris/Yeast

Relatively fast growing cells
Eukaryotic secretion system, post-translational modification (but not quite!)
Often used for secretion
Both transient and stable expression possible

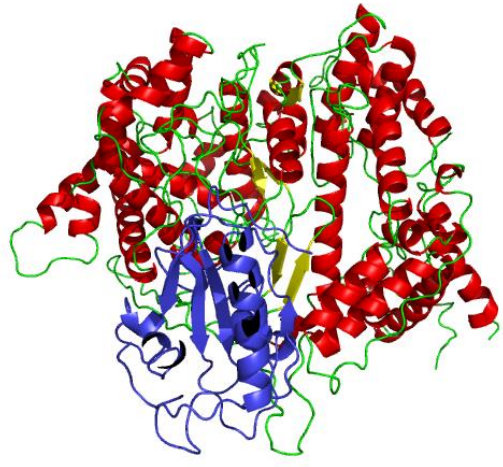


The T7 system

- Promoter of the *gene1* of the bacteriophage T7
- Recognised only by the **T7 RNA polymerase** (T7RP)
  - Faster and more processive enzyme – longer transcripts
- Commercialised in the pET series of vectors from Novagen - tens of variants
- T7RP can be inhibited by T7 lysozyme (pLysS/E plasmids)
- Usually combined with *lacO* regulator and *lacI* gene to provide tighter regulation of expression (T7*lac* promoter)
- Needs to be combined with a T7 transcription terminator (TΦ)

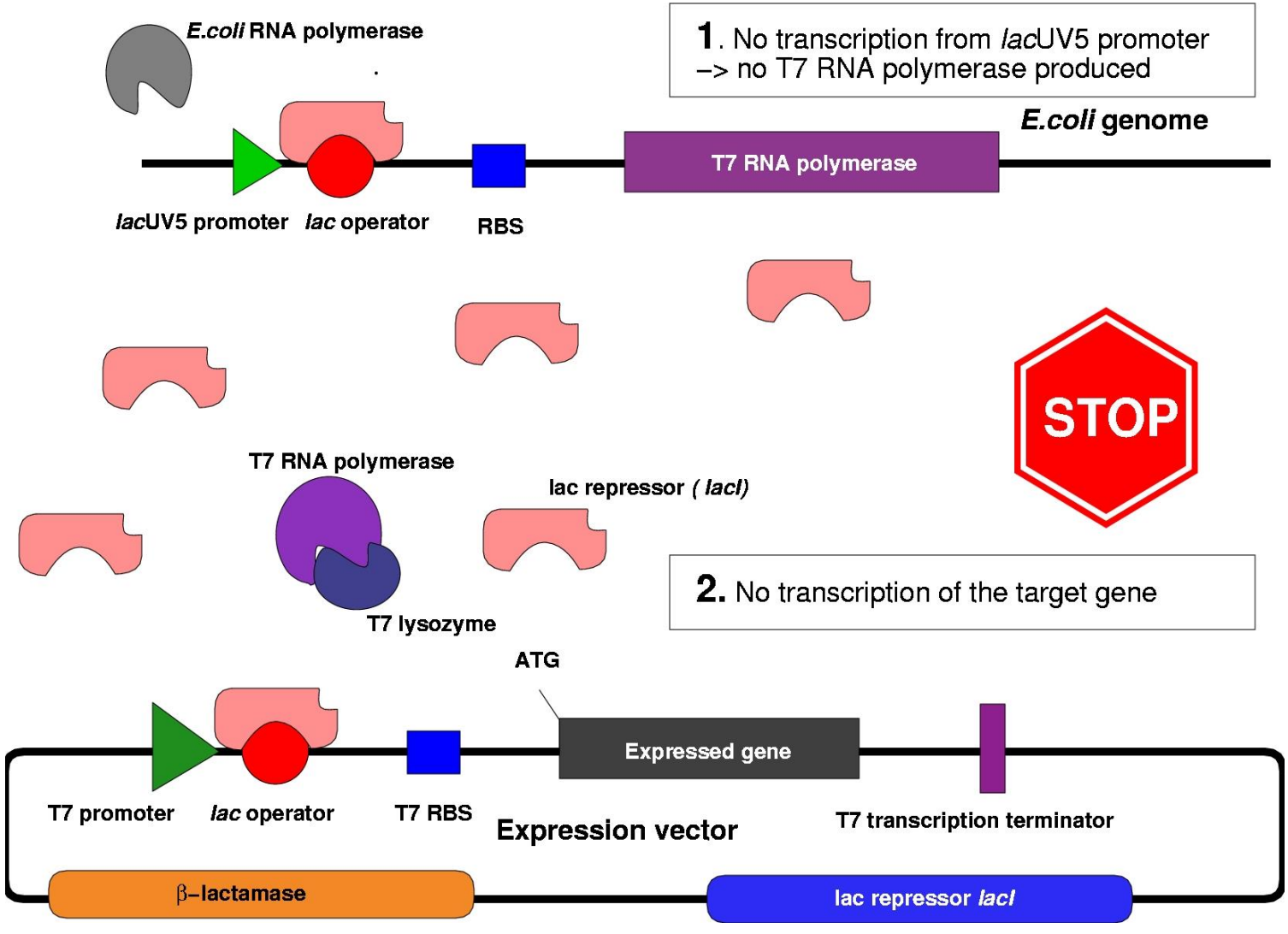


T7RP + promoter



T7RP + T7 lysozyme

T7 system in repressed state

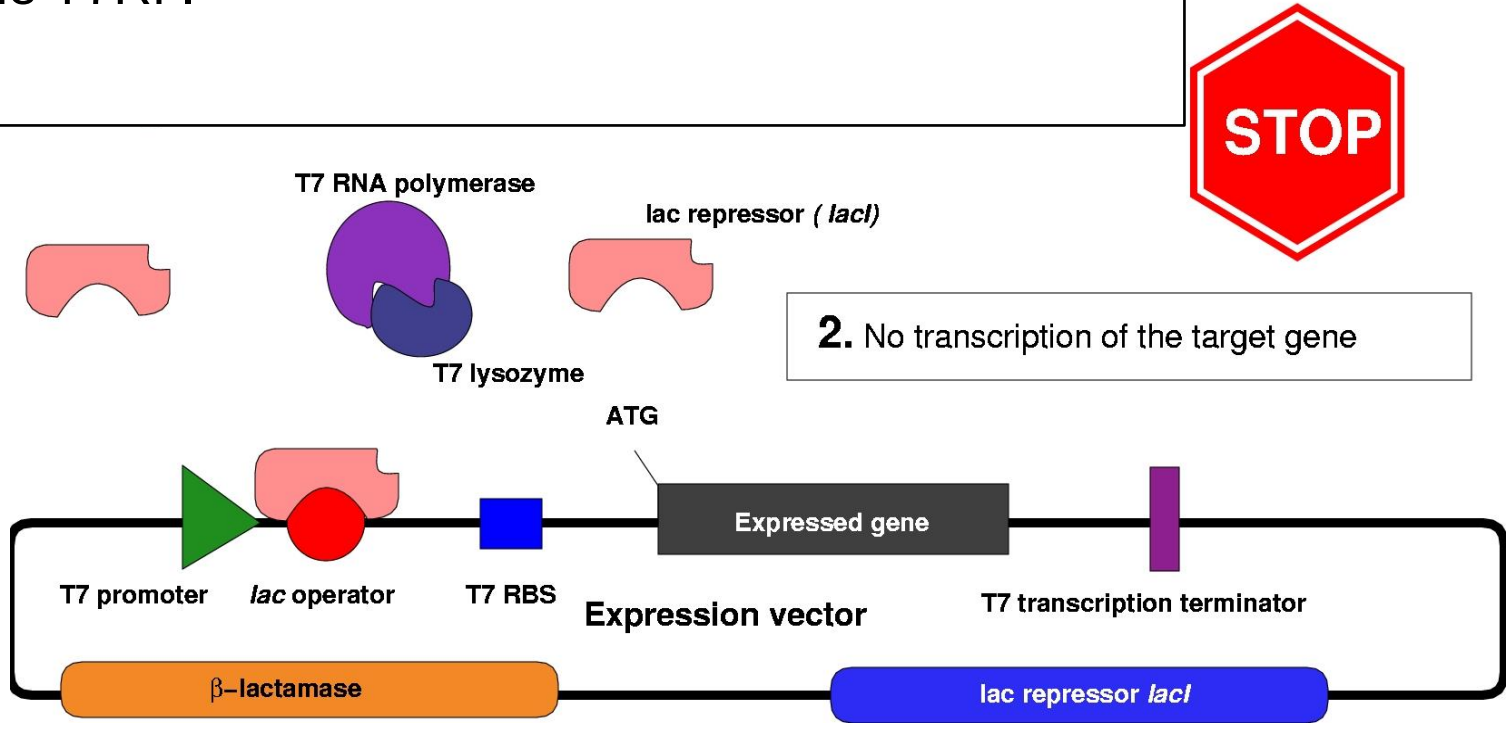


T7 system in repressed state

In a “cloning strain” (DH5α, XL-1Blue...) no T7 RNA polymerase is present and even relatively toxic proteins can be cloned without problems as there should be no expression in the absence of the T7RP.

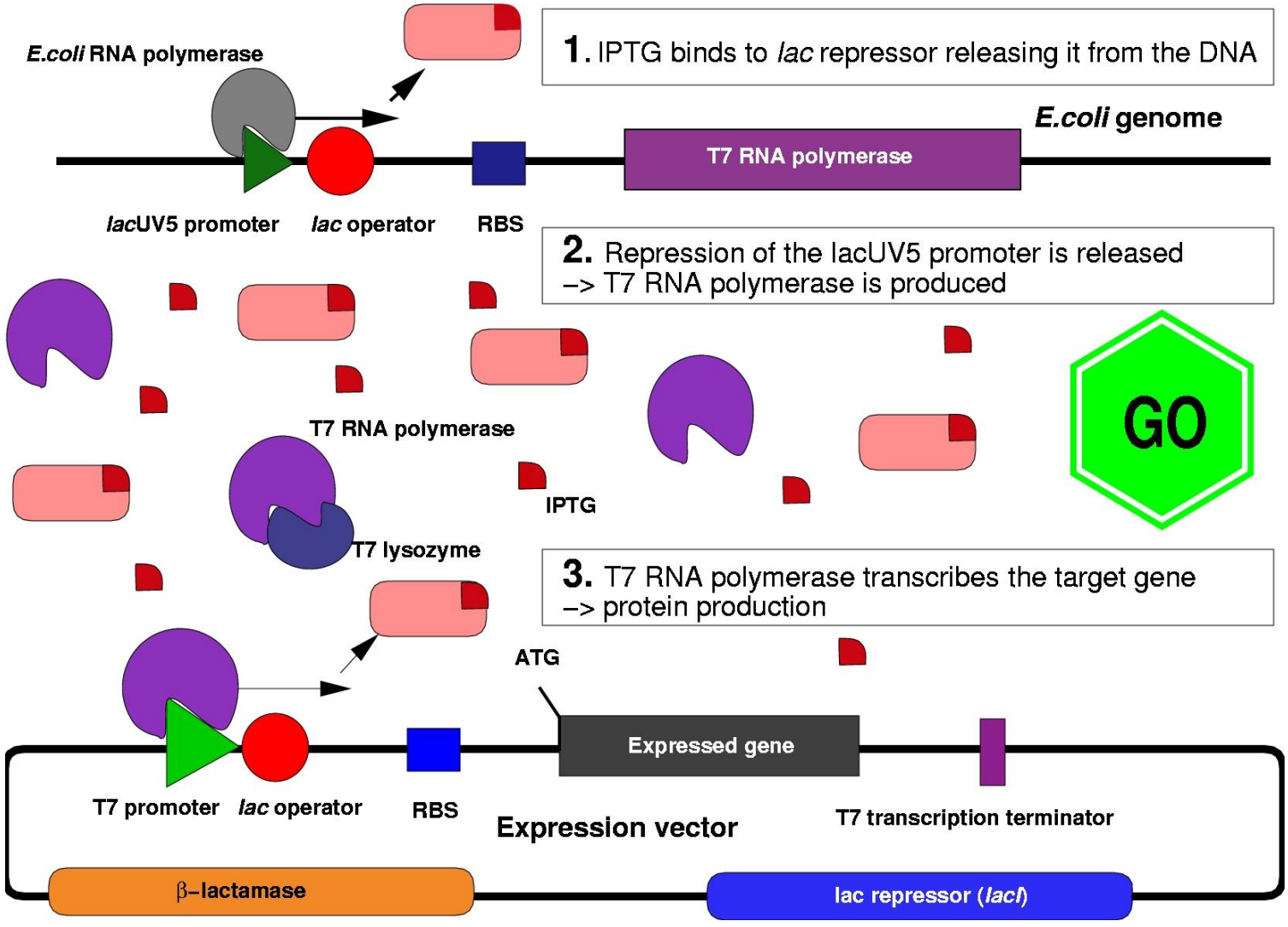
acUV5 promoter
e produced

***E.coli* genome**



2. No transcription of the target gene

Induced T7 system



BL21(DE3): workhorse of T7 system

- Carries lysogenic λ phage (DE3) which contains a copy of the T7 RNA polymerase under the control of *lac*UV5 promoter
- Relatively wild strain, and grows fast (good !)
- A safe, non-pathogenic strain
- Numerous derivatives for special applications
- defective in *OmpT* (outer membrane protease) and, as a B strain, of *lon* protease

More expression strains

Strain	Properties
T7 Express	RecA-, non-phage-derived T7 RNA polymerase. Ok for cloning
Rosetta2(DE3)	Extra copies of rare Arg, Pro, Gly, Leu and Ile tRNA genes
BL21(DE3)RIL (or RP)	As above, but with Arg/Ile/Leu, or Arg/Pro tRNAs
Tuner(DE3)	Lac permease deletion to allow fine-tuning of induction level
BL21(DE3)Star	Mutated RNaseE for RNA stabilisation
BL21-AI	Arabinose inducible T7 RNA polymerase
Lemo21(DE3)	Regulatable T7 RNA polymerase
C41(DE3), C43(DE3)	High expressing, tolerant to membrane protein overexpression
B834(DE3)	Parent strain of BL21(DE3), methionine auxotroph
BL21(DE3)TrxB	trxB deletion strain to facilitate disulfide exchange
OrigamiB(DE3)	trxB/gor deletion for even more oxidising environment
SHuffleT7	trxB/gor deletion and cytoplasmic DsbC co-expression
ClearColi BL21(DE3)	Endotoxin deficient strain for making endotoxin free proteins

Typical expression protocol

Plate cells from glycerol stock, grow overnight — **Day 1, 5pm**

⇩

Scrape all individual colonies and start large scale culture — **Day 2, 9am**

⇩

No need for overnight culture, this is faster and safer

Grown the cells at 37°C until OD_{600} 0.8-1.0 — **Day 2, 2pm**

⇩

If expressing at lower temp, move cells to that temp for 1 h

Induce expression with (typically) 400 μM IPTG

⇩

Express for 3h at 37°C, 4 h at 25°C and overnight at lower temp

⇩

Spin cells, resuspend in MilliQ and freeze — **Day 2, 5pm** / **Day 3, 9am**

Autoinduction with T7 system

- Using a combination of glycerol, glucose and lactose to reach high cell density and **automatic** induction of expression
- Simply grow the cells in a specially formulated medium
  - Once the glucose (which represses the T7*lac* and *lac*UV5 promoters) runs out of the medium, the bacteria will start taking up lactose (note: Tuner(DE3) cells will not work for this!)
  - Lactose induces expression of the target genes slowly, which might also reduce aggregation problems by giving the proteins more time to fold correctly
  - Cells can reach very high culture density, up to OD600 of >20.
- Developed with pET vectors (low/medium copy number) and might need optimisation with other T7 vectors.

Low or no expression

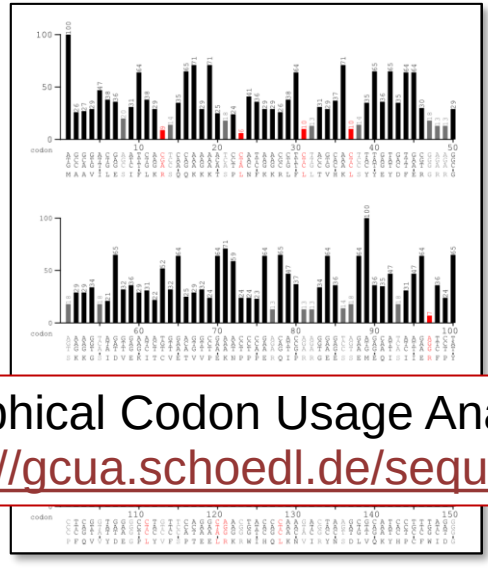
- Are you sure the insert is correct and in frame ?
- Have you checked for rare codons ?
  - Use a strain such as Rosetta(DE3) or Rosetta2(DE3)
  - Codon optimise the gene
- Is the 5′ end of the insert with secondary structure?
  - Optimise the 5′ sequence for reduced GC content
- Is the protein being degraded ?
  - Western blot if antibodies available
- Do you have correct expression controls ?
  - same strain with the same plasmid but without an insert

How overcome codon useage problem?

- Expression in a host which carries extra copies of the limiting tRNA genes.
  - Rosetta(DE3) and Rosetta2(DE3), BL21(DE3)RIL, BL21(DE3)+pUBS520
- Mutate the offending residues
  - Practical only for a few residues
- Synthetic gene with optimised codons
  - The whole coding region synthesised and the codons optimised for particular expression host
  - Can be done as a DIY job, but several companies do it as a service and some even sell them "off the shelf"
  - Getting cheaper – less £100 per construct in some cases

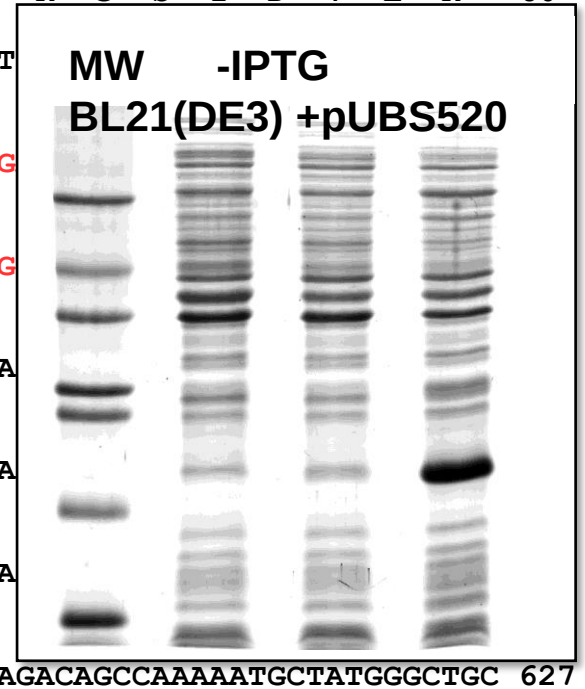
Overcoming rare Arg codon problem

- 170 N-terminal residues of Bruton's tyrosine kinase (Btk)
- 7 AGA/AGGs = 4% of total codons
- 2 tandem pairs
- Expression in BL21(DE3) with or without a plasmid carrying extra copies of the rare tRNA gene
  - pUBS520 plasmid
  - Rosetta, BL21(DE3)RIL strains

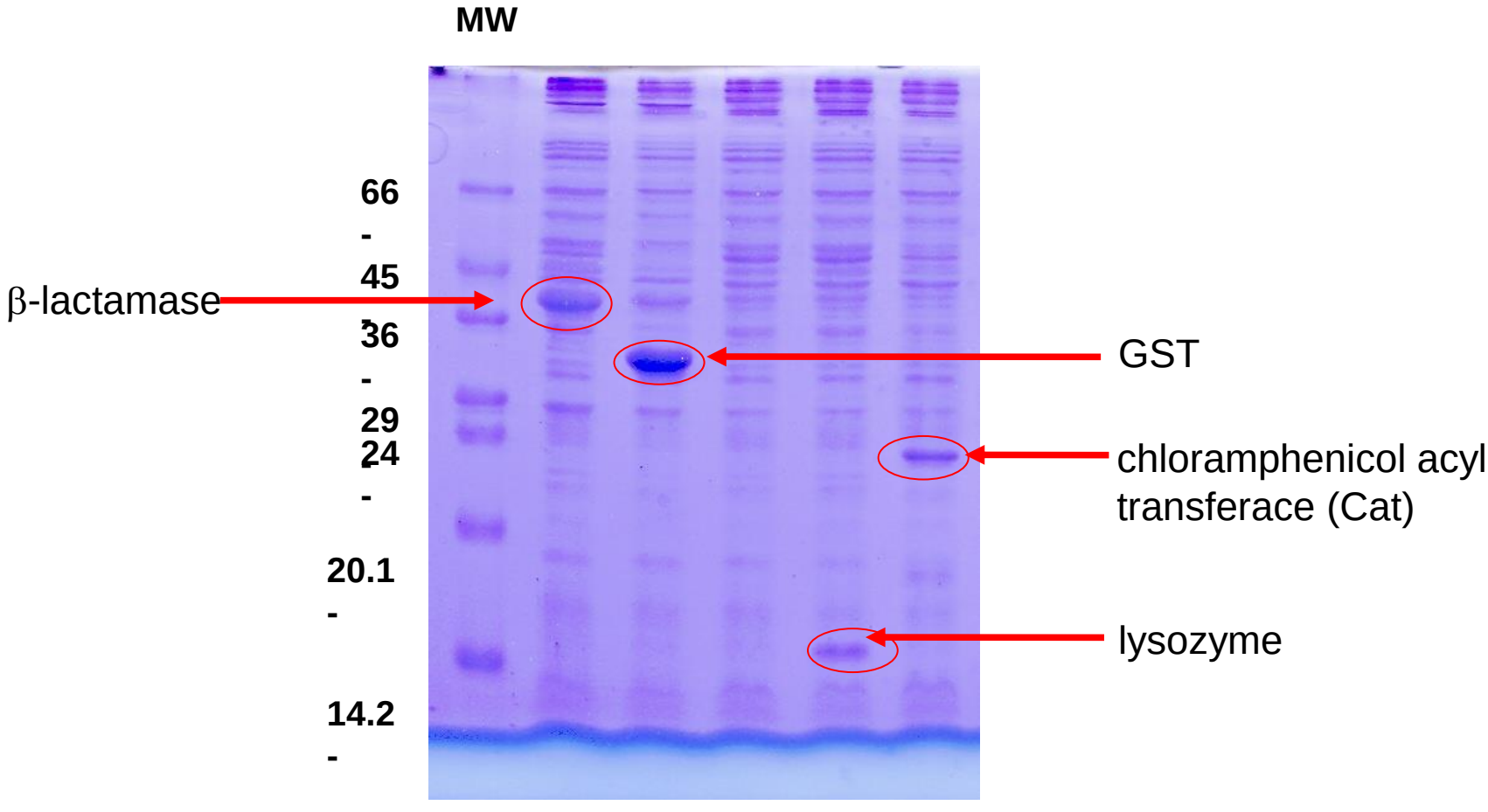


Graphical Codon Usage Analyser
http://gcua.schoedl.de/sequential_v2.html

```
133 ATGGCCGCAGTGATTCTGGAGAGCATCTTTCTGAAGCGATCCCAA 177
  1 M  A  A  V  I  L  E  S  I  F  L  K  R  S  Q   15

178 CAGAAAAAGAAAACATCACCTCTAAACTTCAAGAAGCGCCTGTTT 222
 16 Q  K  K  K  T  S  P  L  N  F  K  K  R  L  F   30

223 CTCTTGACCGTGCACAAACTCTCCTACTATGAGTATGACTTTGAA 267
 31 L  L  T  V  H  K  L  S  Y  Y  E  Y  D  F  E   45

268 CGTGGGAGAAGAGGCAGTAAGAAGGGTTCAATAGATGTTGAGAAG 312
 46 R  G  R  R  G  S  K  K  G  S  I  D  V  E  K   60

313 ATCACTTGTGTTGAAACAGT
 61 I  T  C  V  E  T  V

358 GAAAGACAGATTCCGAGAAG
 76 E  R  Q  I  P  R  R

403 CAAATTTCAATCATTGAAAG
 91 Q  I  S  I  I  E  R

448 TATGATGAAGGGCCTCTCTA
106 Y  D  E  G  P  L  Y

493 AGGAAGCGGTGGATTCACCA
121 R  K  R  W  I  H  Q

538 AGTGATCTGGTTCAGAAATA
136 S  D  L  V  Q  K  Y

583 CAGTATCTCTGCTGCTCTCAGACAGCCAAAAATGCTATGGGCTGC 627
151 Q  Y  L  C  C  S  Q  T  A  K  N  A  M  G  C  165
```

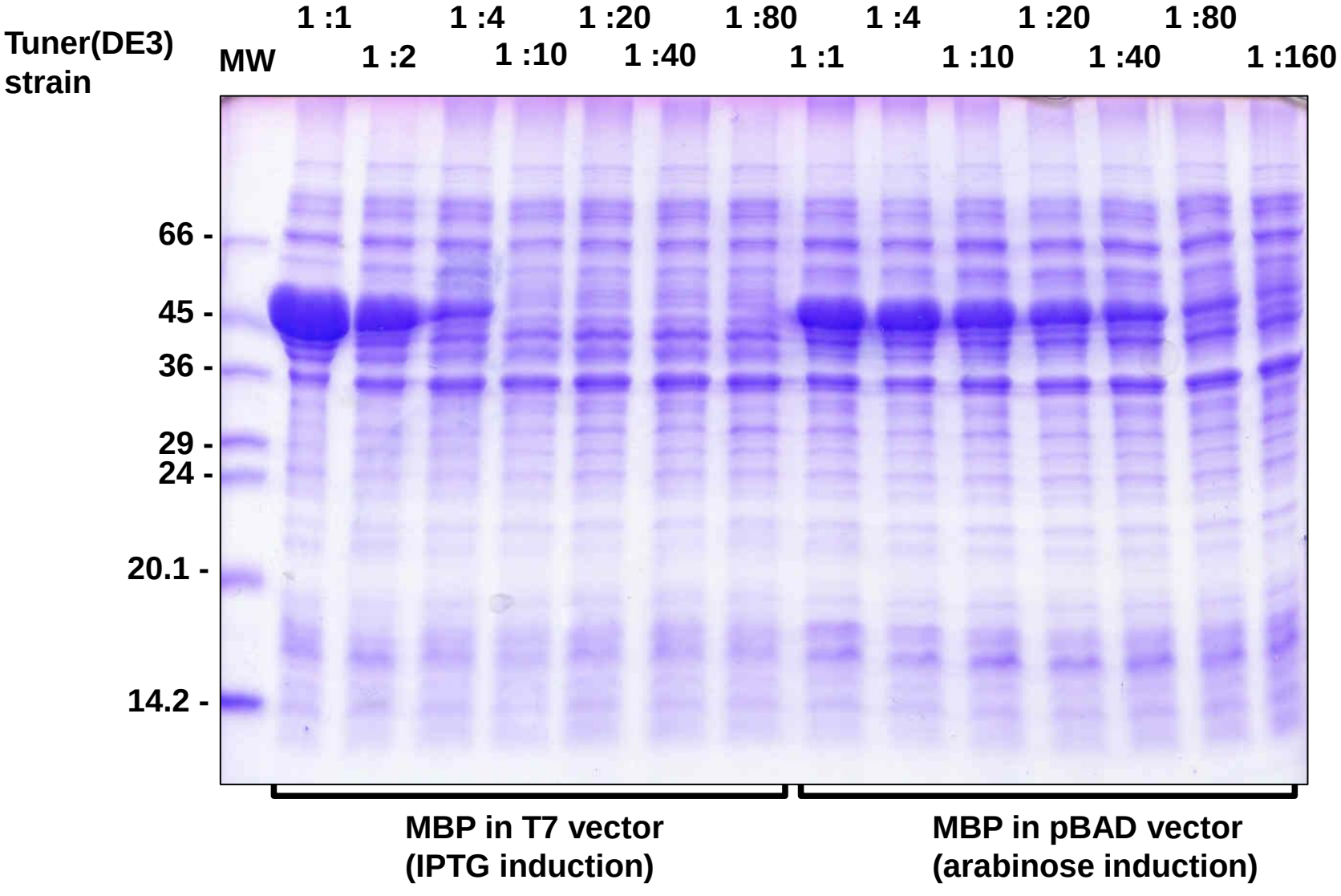


Expression testing: Where is my protein ?



Correct negative control is the same expression plasmid withoutan insert, grown and induced like the target plasmid. And it needs to be in each gel!

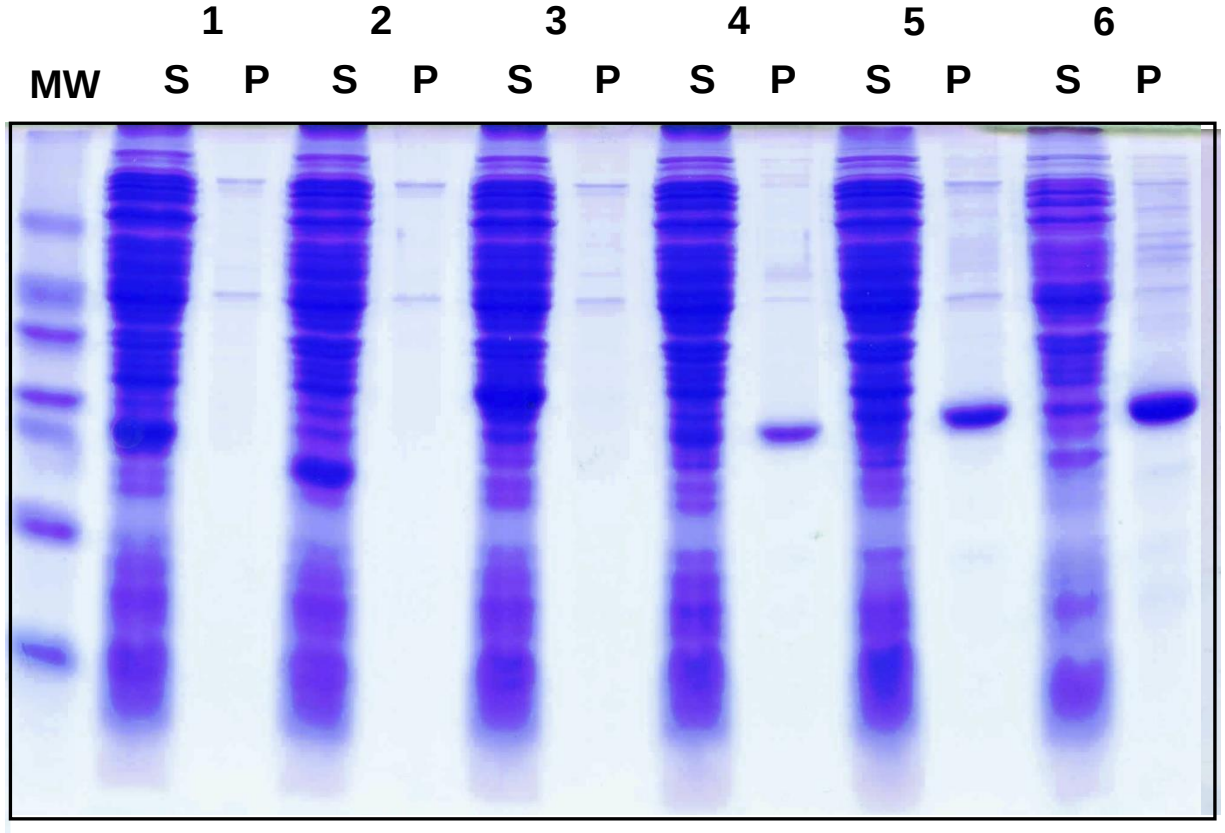
Regulating the expression level by partial induction



Insolubility

- Very common and frustrating problem with *E. coli*
- Formation of inclusion bodies (IBs) can occur even with natural *E. coli* proteins (such as β-galactosidase)
- Their formation is a combination of many effects, namely
  - high speed and level of expression
- Best to try lowered expression temperature
  - down to 15°C, possibly combined with partial induction
- But….
  - Protein IBs are protected from proteases
  - The protein is already relatively pure
  - Can you refold it? (lecture to come on that !)

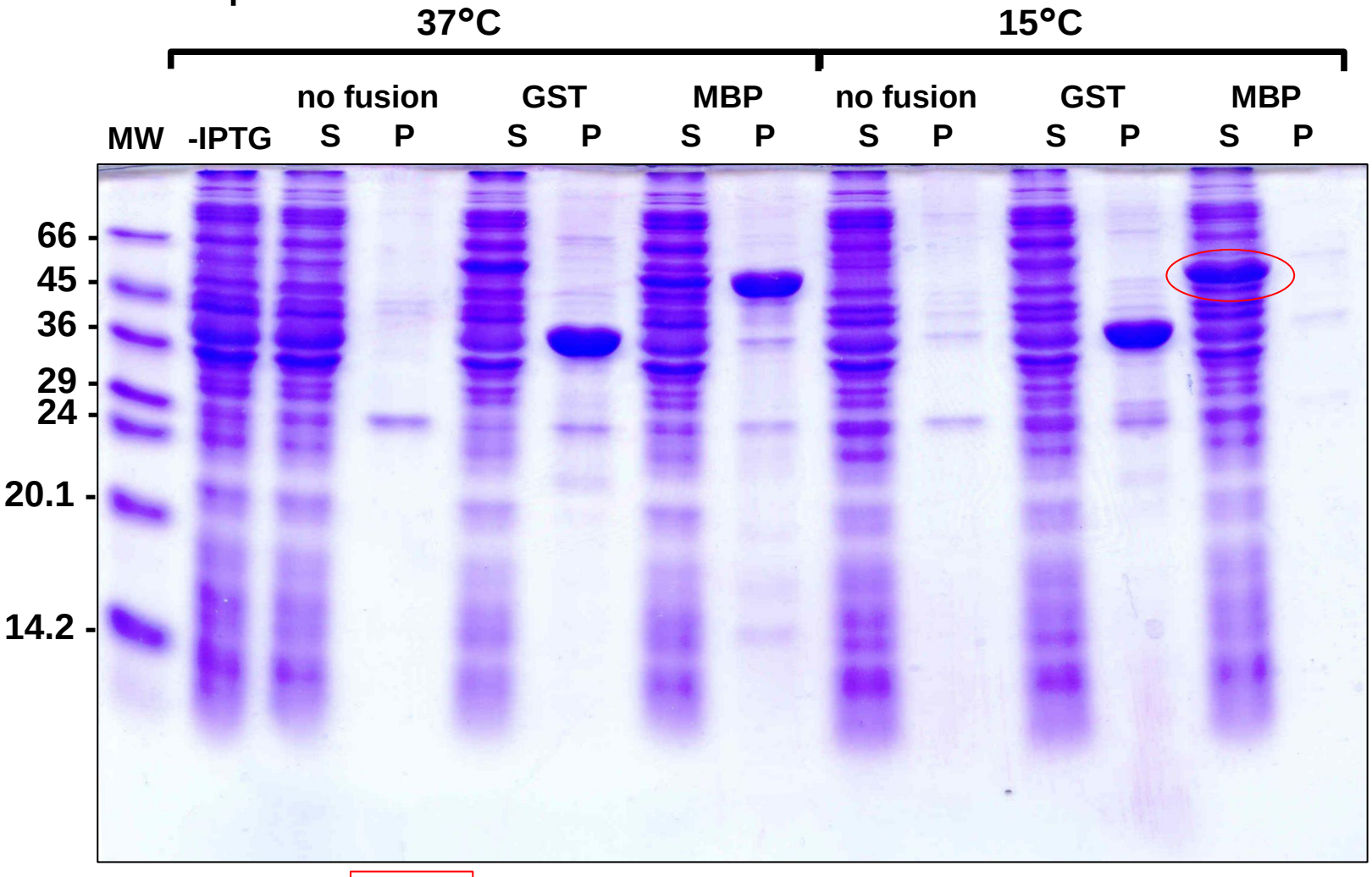
Testing for solubility

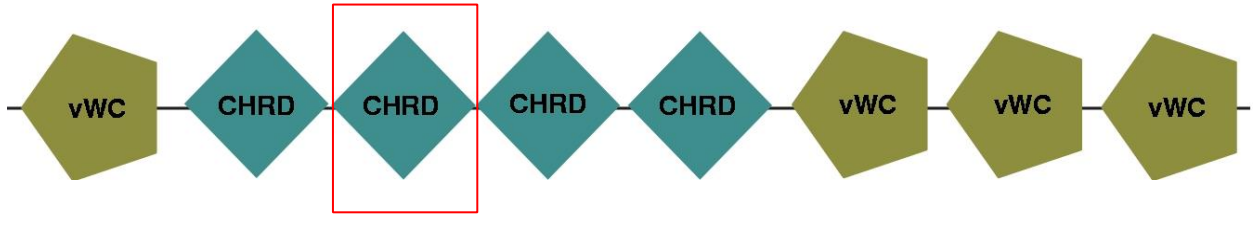


Cells need to be properly lysed:

- in small scale we use BugBuster *with* lysozyme and DNase to ensure complete lysis.
- Do not forget those all important controls !

Fusion as a solubility enhancer: expression of CHRD2 domain





Chordin

Effect of *E.coli* strain and fusion: expression of trxA-AR2B