Expression systems

1. Addendum to last lecture: star activity
2. Different expression systems
3. Questions related to expression system-specifc plasmids
4. The cloning challenge

Star activity

5'-GTTGG AATTCAGGA-3'
3'-CAACCTTAA GTCCT-5'

EcoRI

5'-GTTGC AATTCAGGA-3'
3'-CAACGTTAA GTCCT-5'

5'-GTTGT AATTCAGGA-3'
3'-CAACATTAA GTCCT-5'

EcoRI*

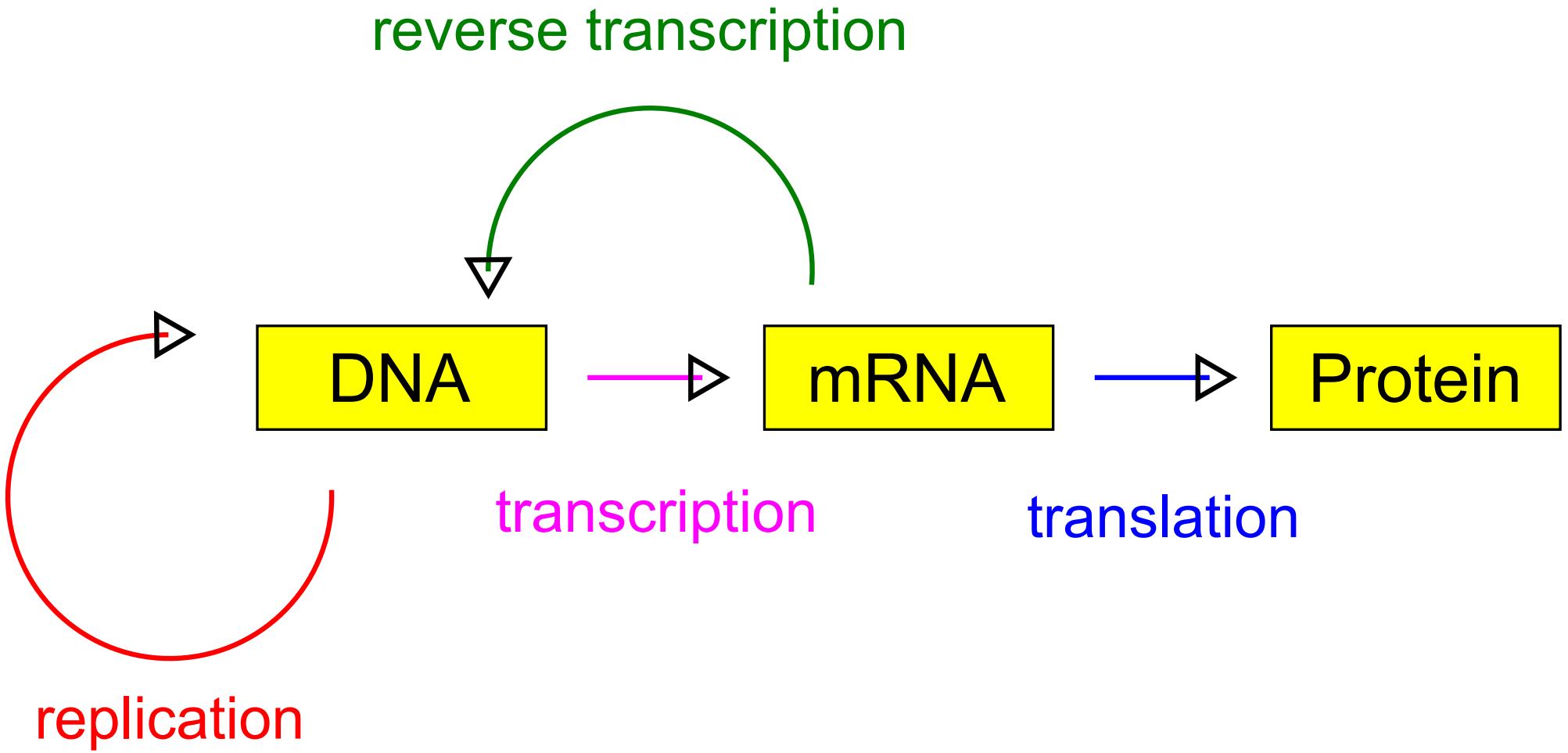
- High glycerol concentration
- Wrong buffer (e.g. too low salt concentration)
- Too much enzyme, too long incubation

→ **High-Fidelity enzymes**

Science or Fiction

- Craig Venter's 1st synthetic cell had neither synthetic cytoplasm nor synthetic DNA, only the DNA fragments that were assembled by a yeast cell in a multistep process were synthetic.
- The NIH originally wanted to confine all recombinant DNA technology to class 3 facilities.
- A company has produced a DNA laser printer that uses lasers to select error free oligonucleotides for assembly into contigous DNA streches.
- The first type II restriction enzyme HindIII was discovered in 1959 by Arber and Meselson.

Central Dogma of Molecular Biology

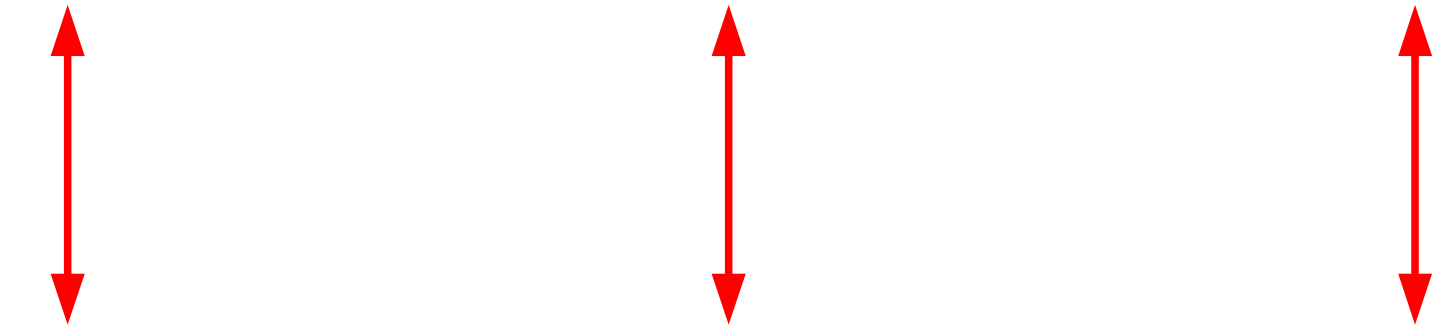


Different methods

- Stable expression
- Transient expression

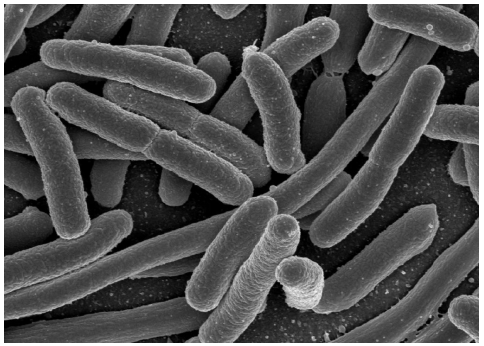
- Inducible expression
- Constitutive expression

- Nonviral delivery
- Viral delivery

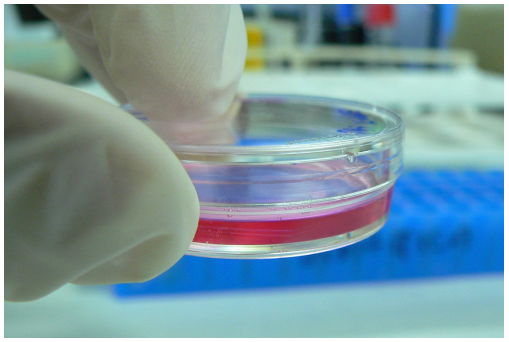


Different expression hosts

- **Bacteria**
  E.coli, Bacillus subtilis, Lactococcus, Pseudomonas

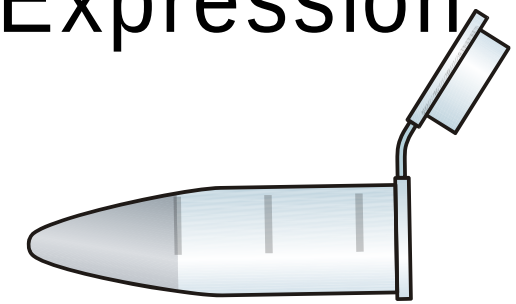


- **Mammalian cells**
  HEK293, CHO

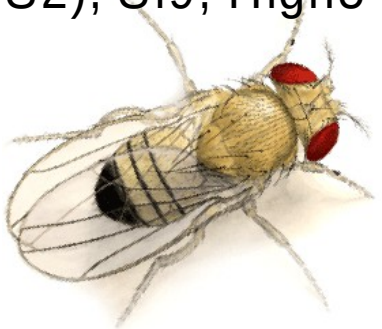


- **Yeast**
  Saccharomyces cerevisiae, Pichia pastoris

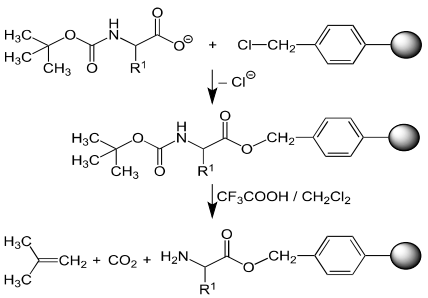
- **Cell-free Expression**



- **Insect cells**
  Drosophila (Schneider S2), Sf9, High5



- **Chemical Protein Synthesis**

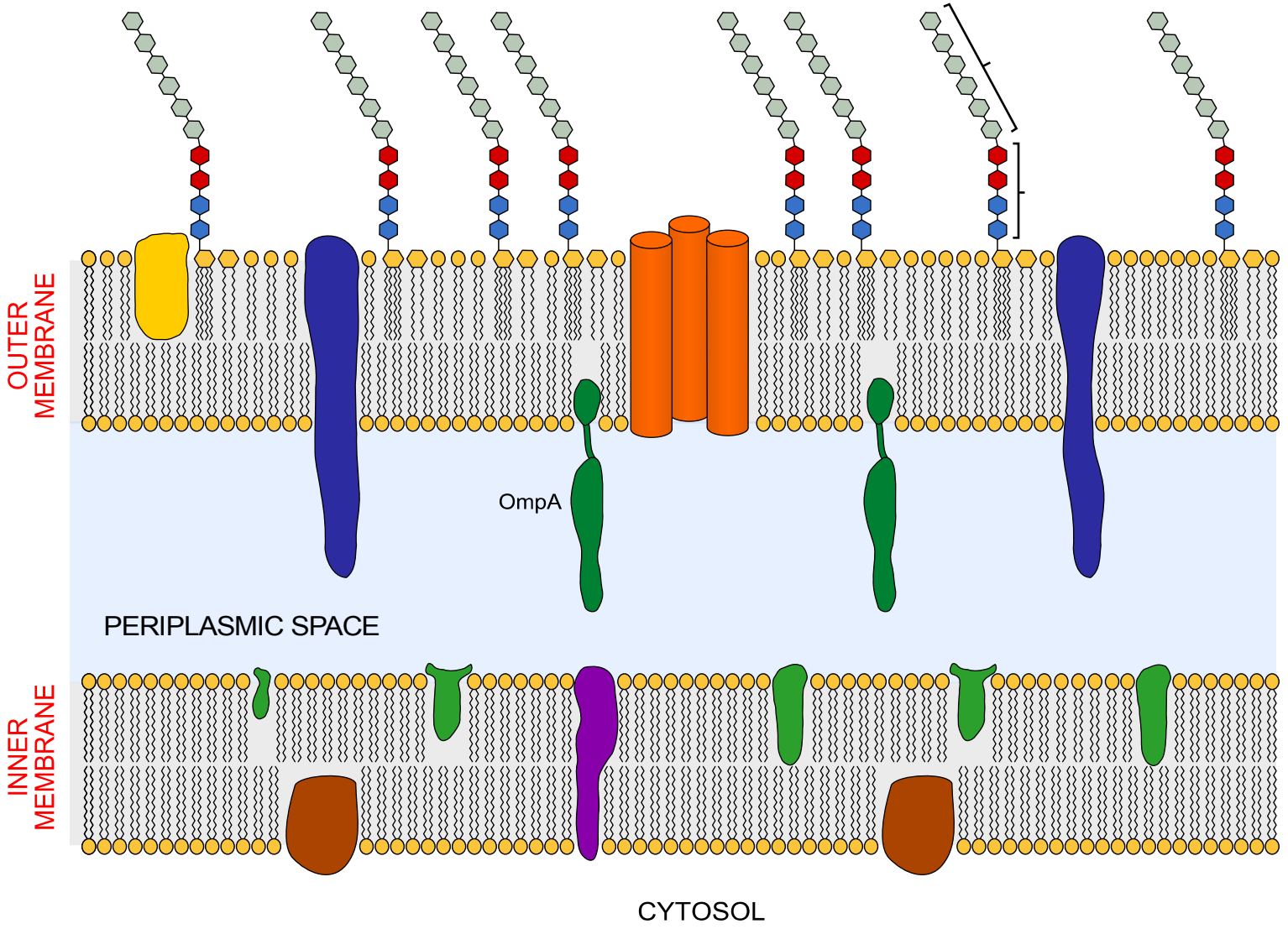


Bacterial expression

- Cheap!
- Fast!
- Several mammalian proteins can be expressed easily in E.coli!
- Sometimes you don't need >90% activity, but 5% might be enough (IL-3)
- Several commercial systems, e.g. pET, pBAD, GST
- Secreted expression into periplasm
- The first biopharmaceutical made with recombinant DNA technology was E.coli-produced human insulin, still some of the insulins are produced in E.coli (e.g. the inhalable insulin)
- Distribution of all biopharmaceuticals by expression system (2011):

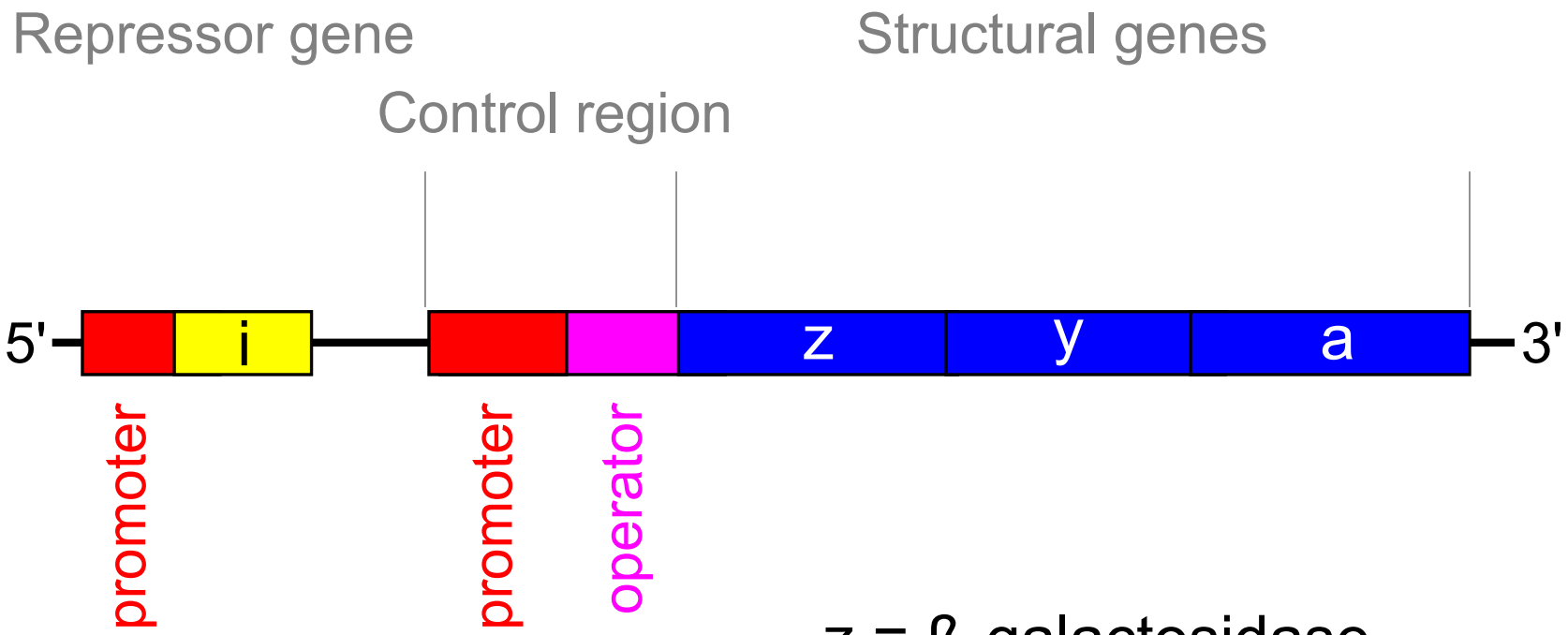
	E.coli	Yeast	Mammalian cells
%	31	15	43

Secreted protein expression in E.coli



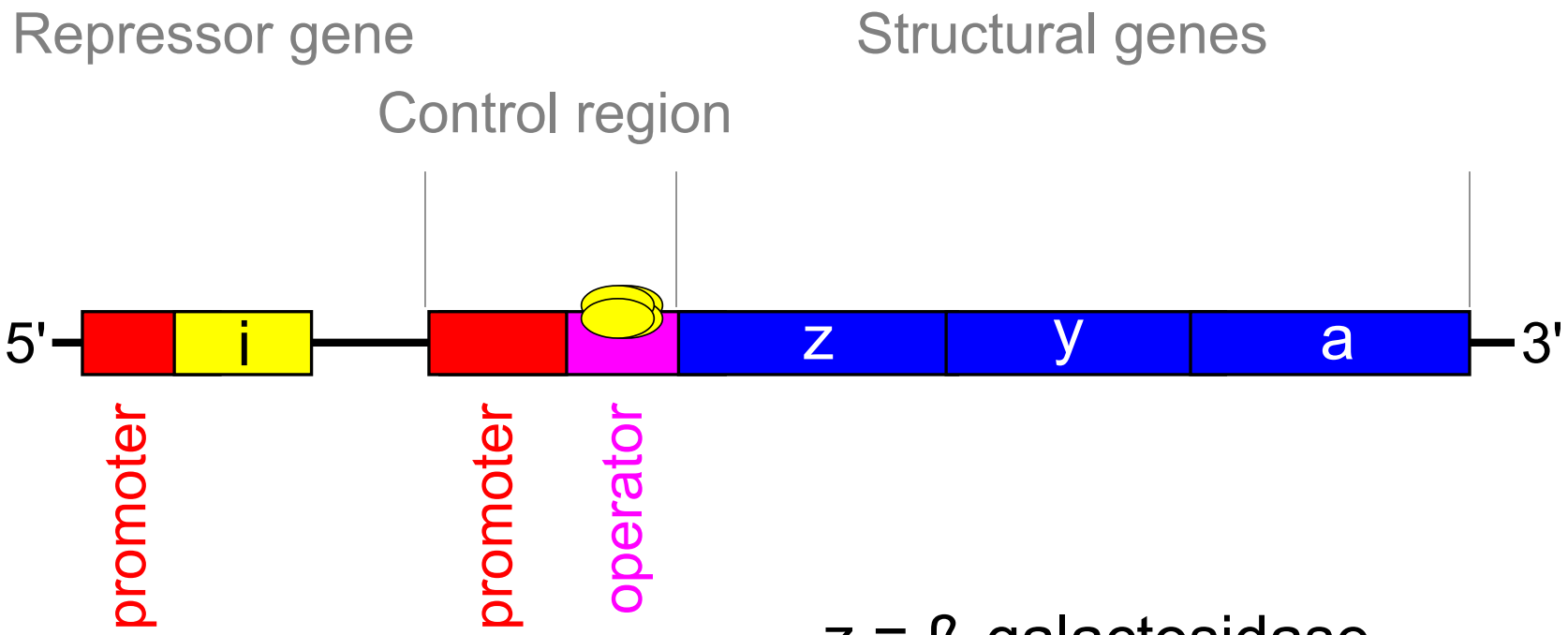
Modified from: *Gram negative cell wall* by Jeff Dahl. Creative Commons License.

The lac operon



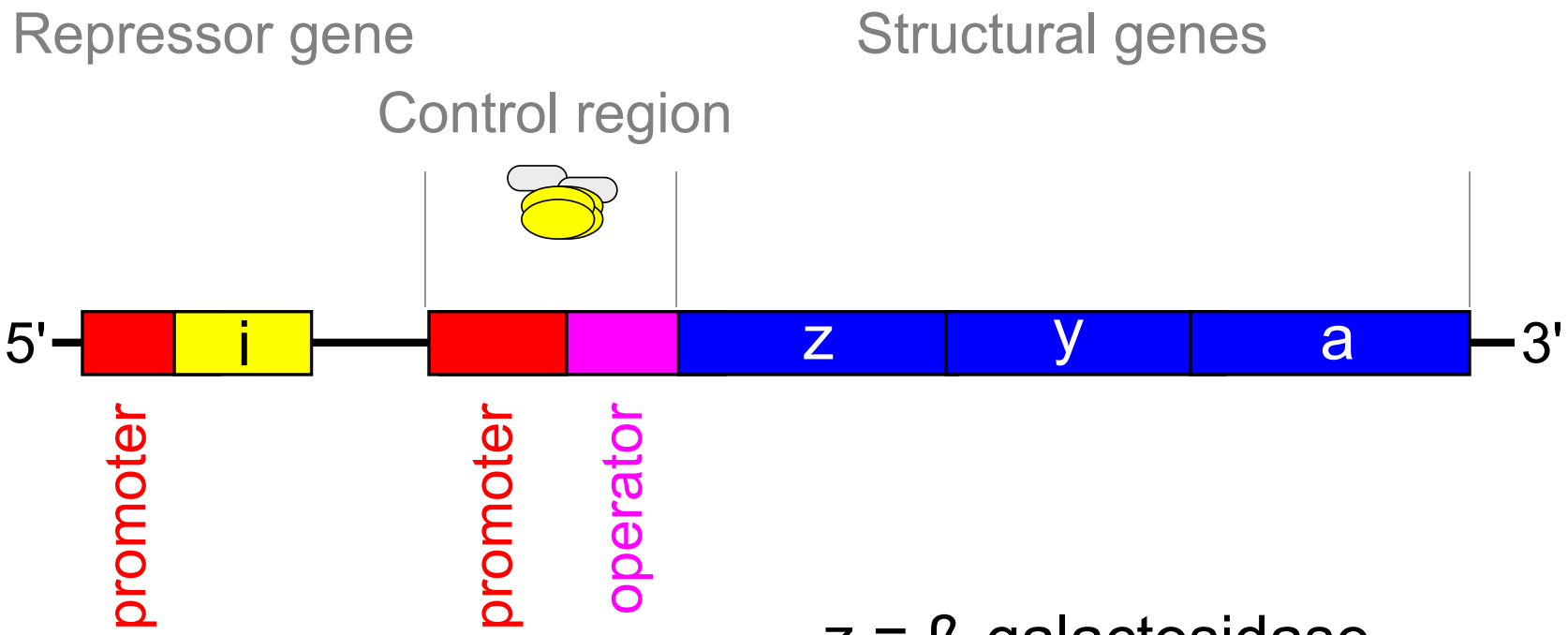
z = ß-galactosidase
y = galactoside permease
a = transacetylase
i = lac repressor

The lac operon



z = ß-galactosidase
y = galactoside permease
a = transacetylase
i = lac repressor

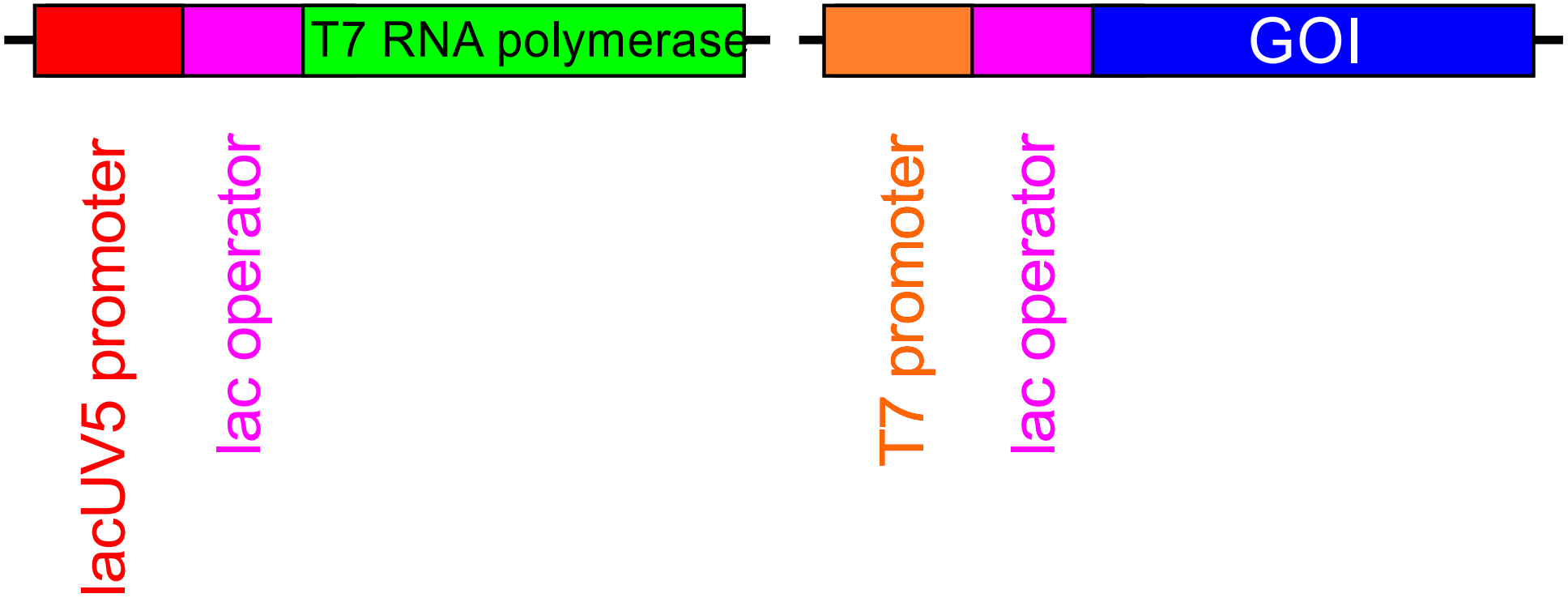
The lac operon



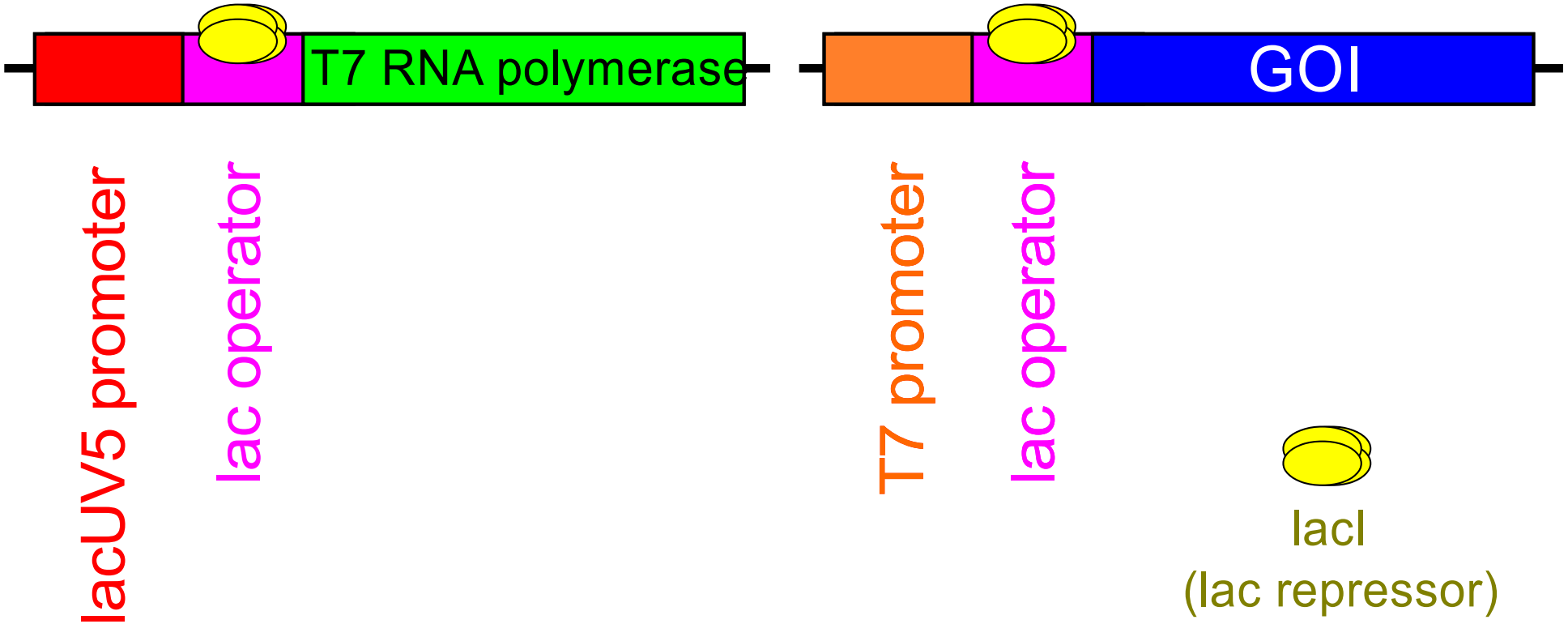
z = ß-galactosidase
y = galactoside permease
a = transacetylase
i = lac repressor

Mostly relatively short promoters are used (Amp 105bp, EM7 48bp, T7 19bp)
No polyA, but other termination signals (GC-rich palindromes + polyT stretch)

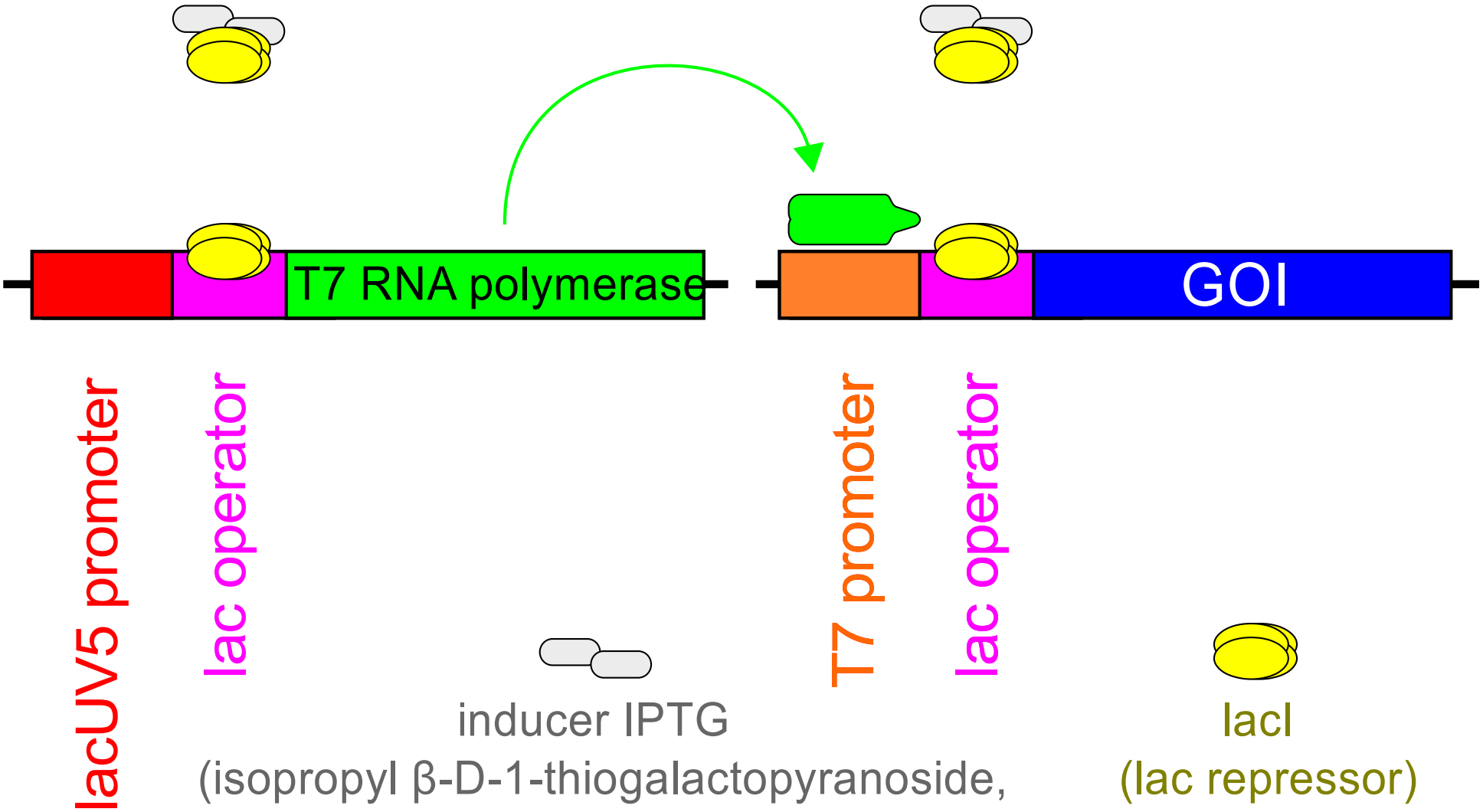
Bacterial expression in E.coli



Bacterial expression in E.coli



Bacterial expression in E.coli



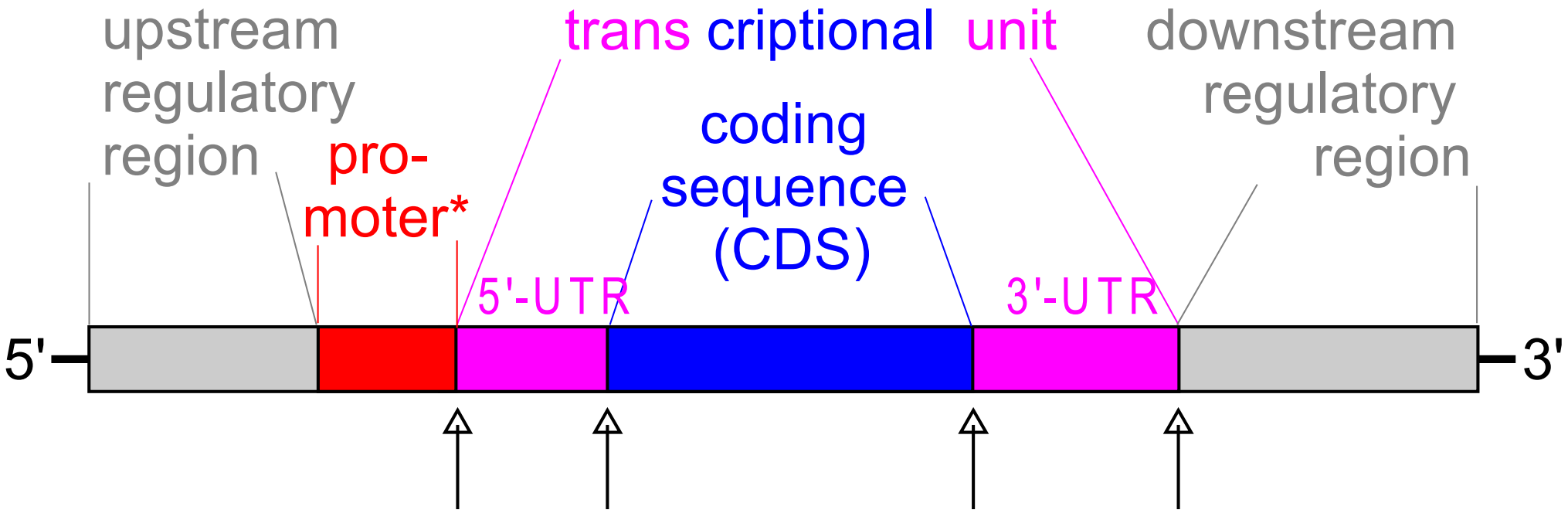
Bacterial expression in E.coli

- Special E.coli strains: e.g. BL21(DE3) or BL21(DE3)pLysS
  DE3 = T7 RNA polymerase under lacUV5/lacO control (genomic)
  pLysS = T7 lysozyme, which inhibits T7 RNA polymerase
- Many other systems with tighter regulation
- Constitutive expression: EM7 promoter for non-toxic proteins
- GOI up to 50% of total cell protein
- Solubility: inclusion bodies (partially predictable, main factor: charge)
- Increasing the solubility: MBP (maltose binding protein)- or GST (gluthathion-S-transferase)-tag; low temperature growth
- Has anybody else managed to express your GOI successfully in E.coli (including commerical suppliers)?
- Inclusion bodies can be useful for purification if they can be solubilized and refolded!

Yeast

- Can be sometimes an alternative...
- VEGF-C produced in yeast is active, while its closest homolog VEGF-D is inactive (glycosylation issue)

Eukaryotic gene organization



transcriptional start site

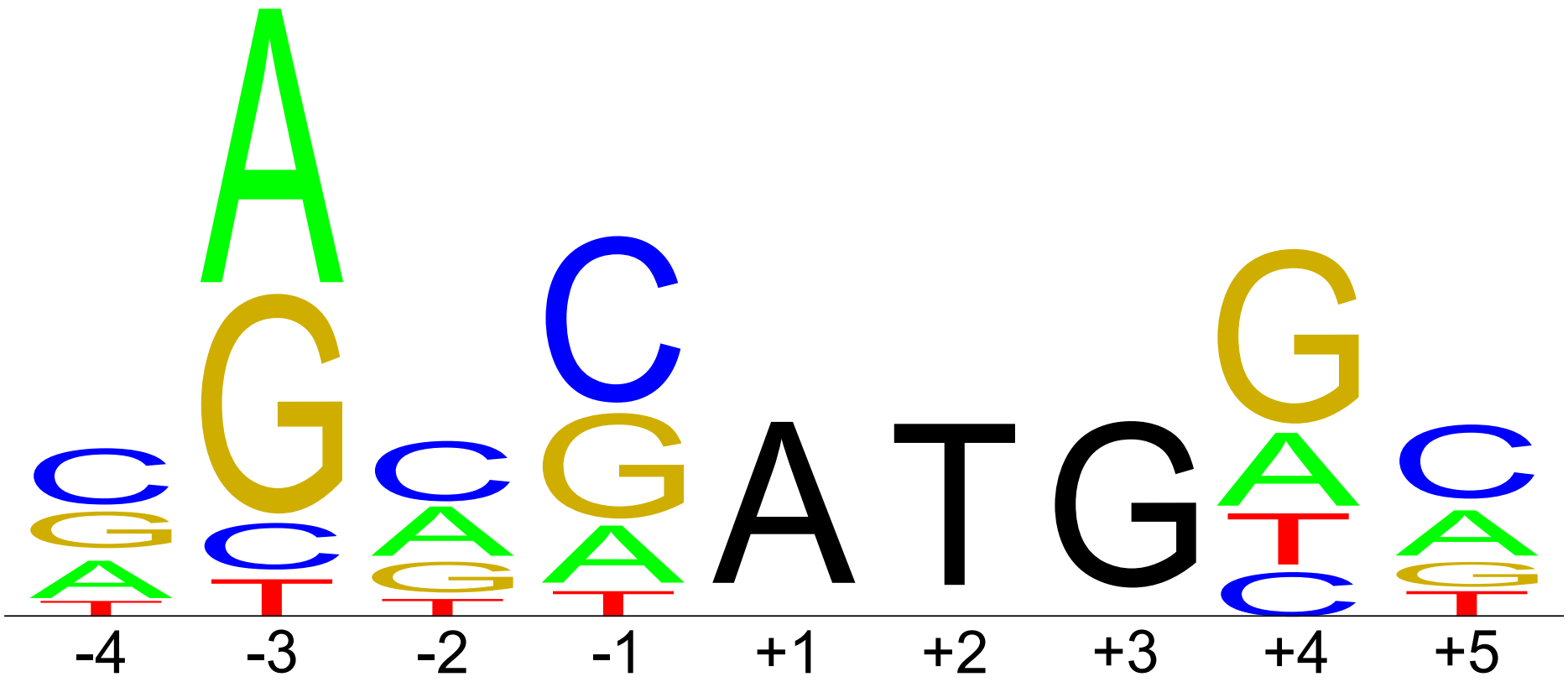
trans-lational start site ("Kozak+ATG")

trans-lational stop site ("stop codon")

transcriptional stop site ("polyA-signal")

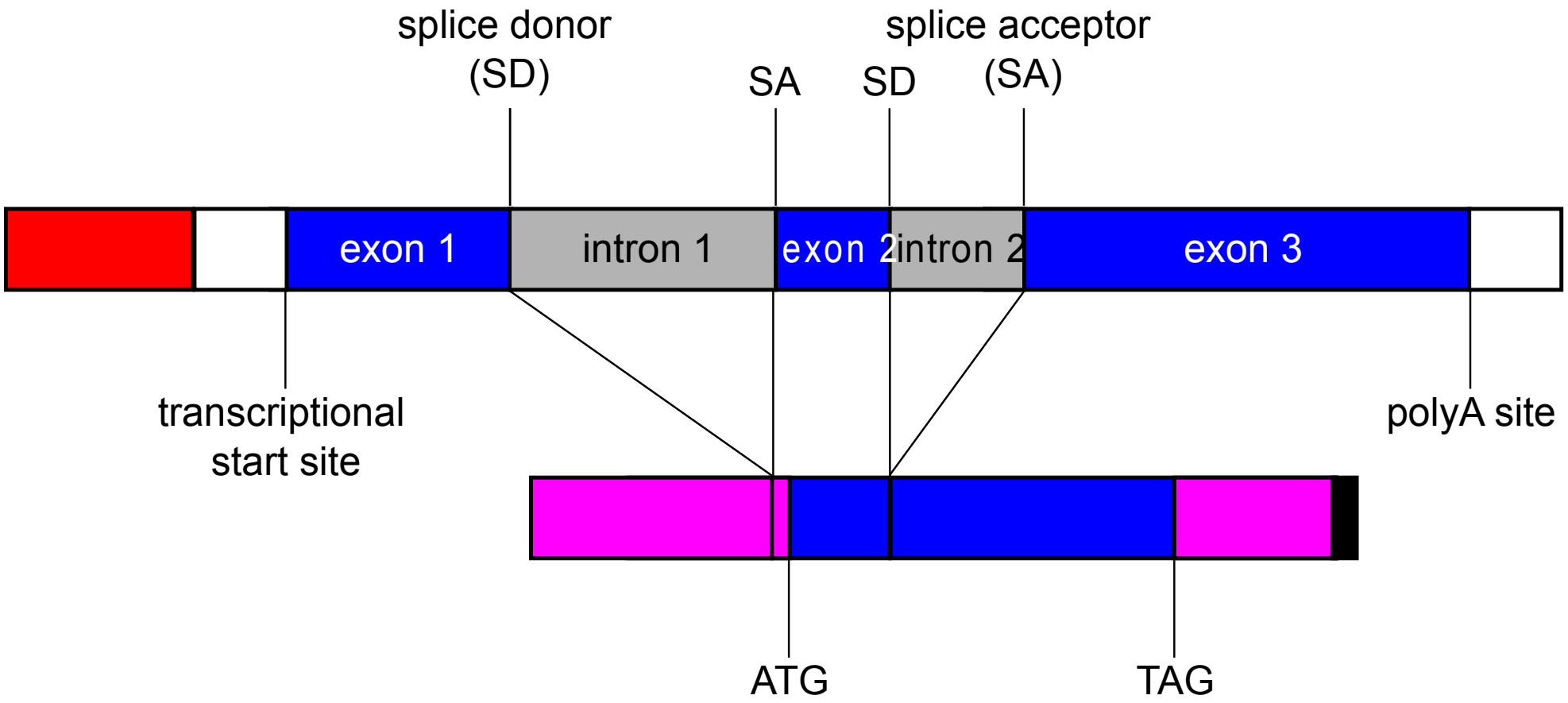
* = RNA polymerase binding site

Kozak...



- For eukaryotic initiation!
- Kozak used mammalian sequences! Insect cells, yeast, etc. have divergent initiation preferences (although the purin preference at -3 seems universal)

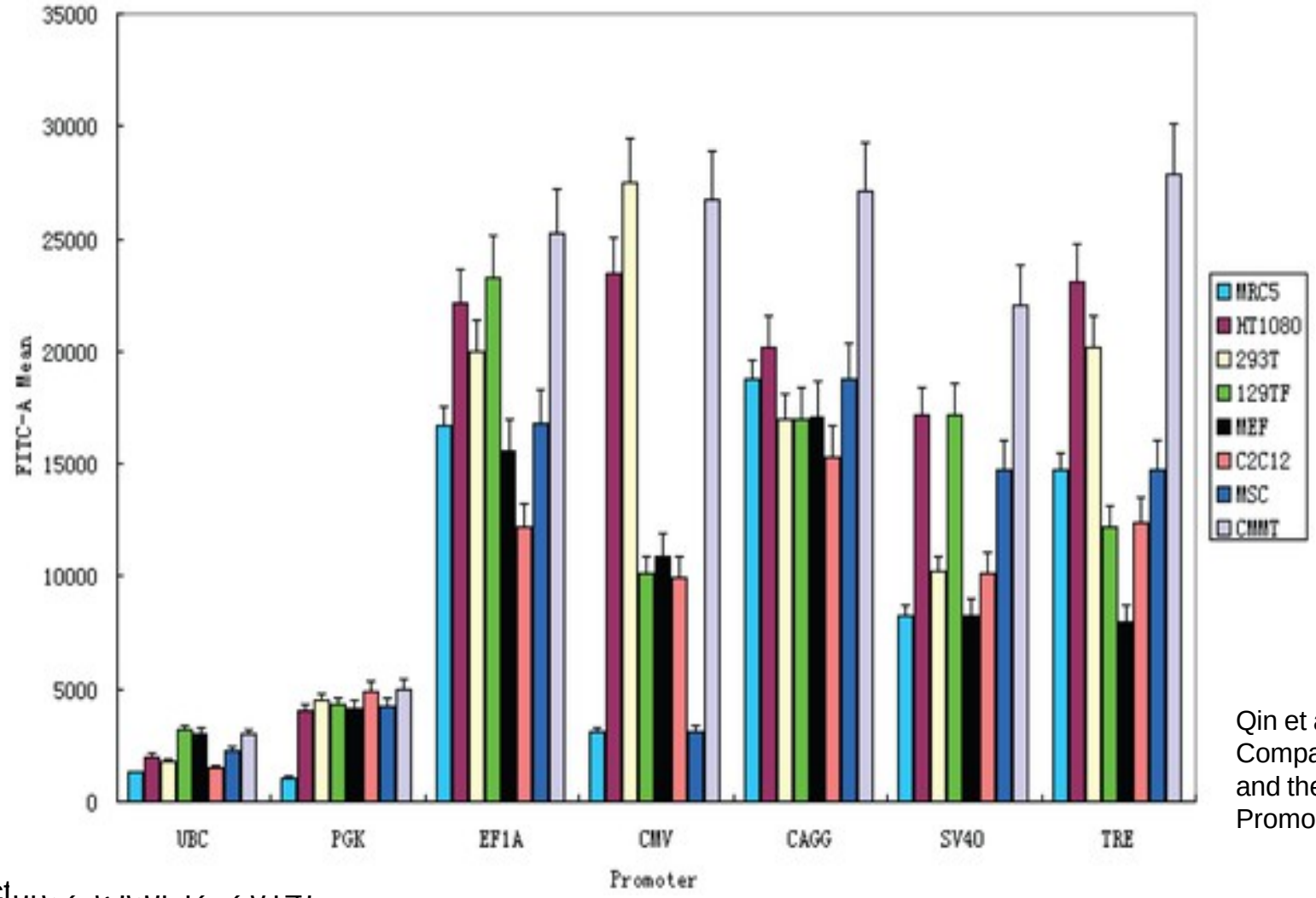
To intron or not to intron



- Introns can increase expression levels and are included in some vectors (typically before the MCS)
- Splice donor and splice acceptor sites are weakly defined, risk to activate „cryptic“ splice sites or to generate them de-novo
- Can be checked, e.g. at http://spliceport.cbcb.umd.edu/

Promoter choice

- Constitutive vs. inducible promoters
- Promoter are species and cell-line specific (due to specific transcription factor binding sites)
- To bypass this restriction, mostly viral promoters (which use general and virally encoded transcription factors) are used that function in most cell lines



Qin et al. (2010) Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. PLoS ONE 5(5): e10611.

polyA signal

- Important for mRNA export from nucleus, mRNA stability and recruitment of ribosomes
- Not as species-specific like promoters
- In bacteria, polyadenylation promotes mRNA degradation.
- No absolute requirement

Insect cells

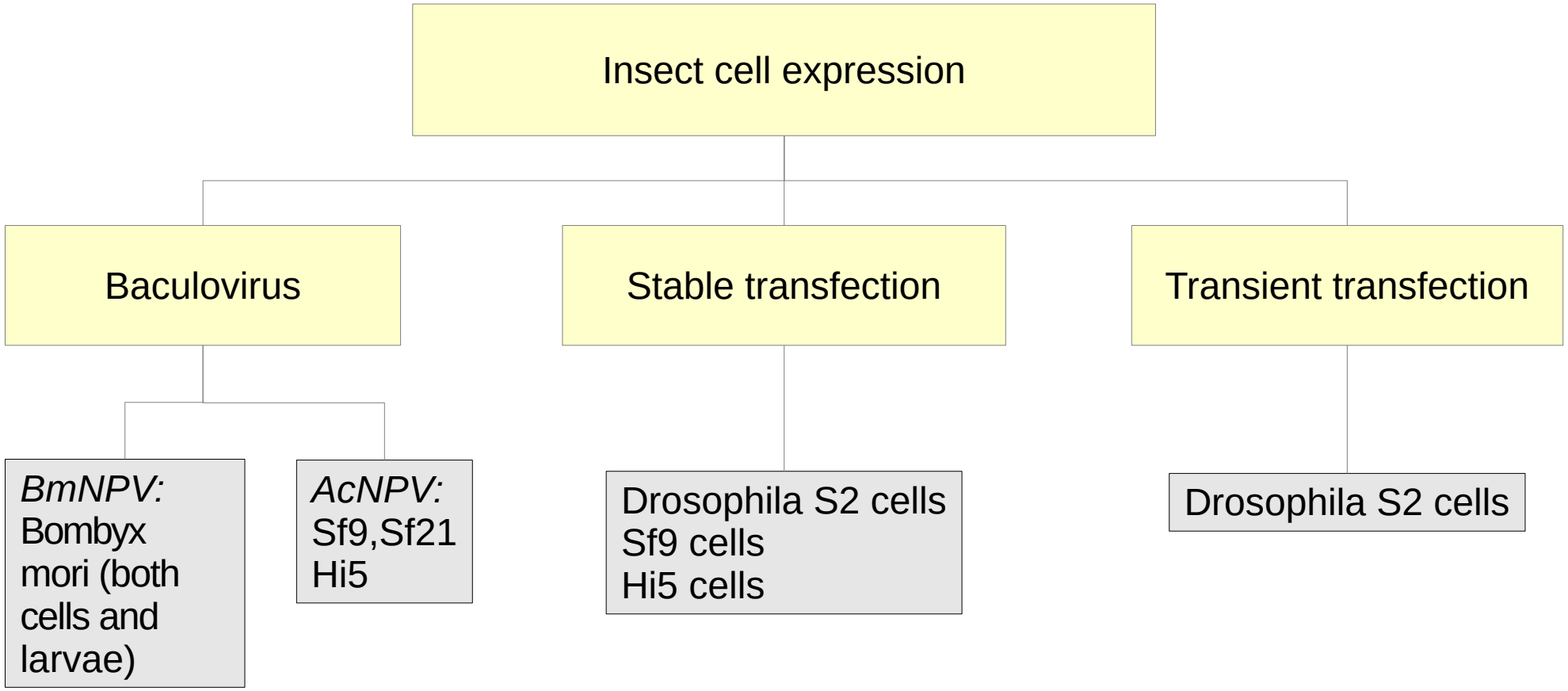
- Very few insect cell lines:
  Schneider S2
  Sf9, Sf21 (Spodoptera frugiperda)
  Hi5 (BTI-TN-5B1-4, Trichoplusia ni)

Drosophila melanogaster by André Karwath

Spodoptera frugiperda by Canadian Biodiversity Information Facility

Trichoplusia ni larva by Alton N. Sparks, Jr.

Common insect cell expression systems



Bac-to-Bac®*

DES*
InsectSelect™*

*Invitrogen/LifeTechnologies

Good to remember...

- Mammalian promoters (CMV, SV40, EF1α) do not work in insect cells! Special insect cell plasmids are needed. Baculoviral promoters do not work in S2 cells and S2 cell promoters do not work in Sf or Hi5 cells!
- For secreted proteins: Mammalian signal peptides may or may not work!
- Baculoviruses require class 1 biosafety level (they do efficiently infect mammalian cells, hence they are considered as gene therapy vectors)
- Typically 3-4 weeks from ready construct to protein expression
- Transfection of insect cells is not as straightforward as transfection of mammalian cells; fewer well-performing transfection reagents on the market than for mammalian cells (e.g. Qiagen's Effecten)
- Insect cells don't need CO_2-incubators and can be grown at RT!
- Selection of stable cell lines is usually done without clonal isolation.

Mammalian expression systems

- Mostly stable cell lines are used:
  CHO (Chinese Hamster Ovary) cells, ~90% of all biopharmaceuticals
  293 HEK (Human Embryonic Kidney)
- Transient transfection works as well (and is faster, but costs for transfection reagents become prohibitive)

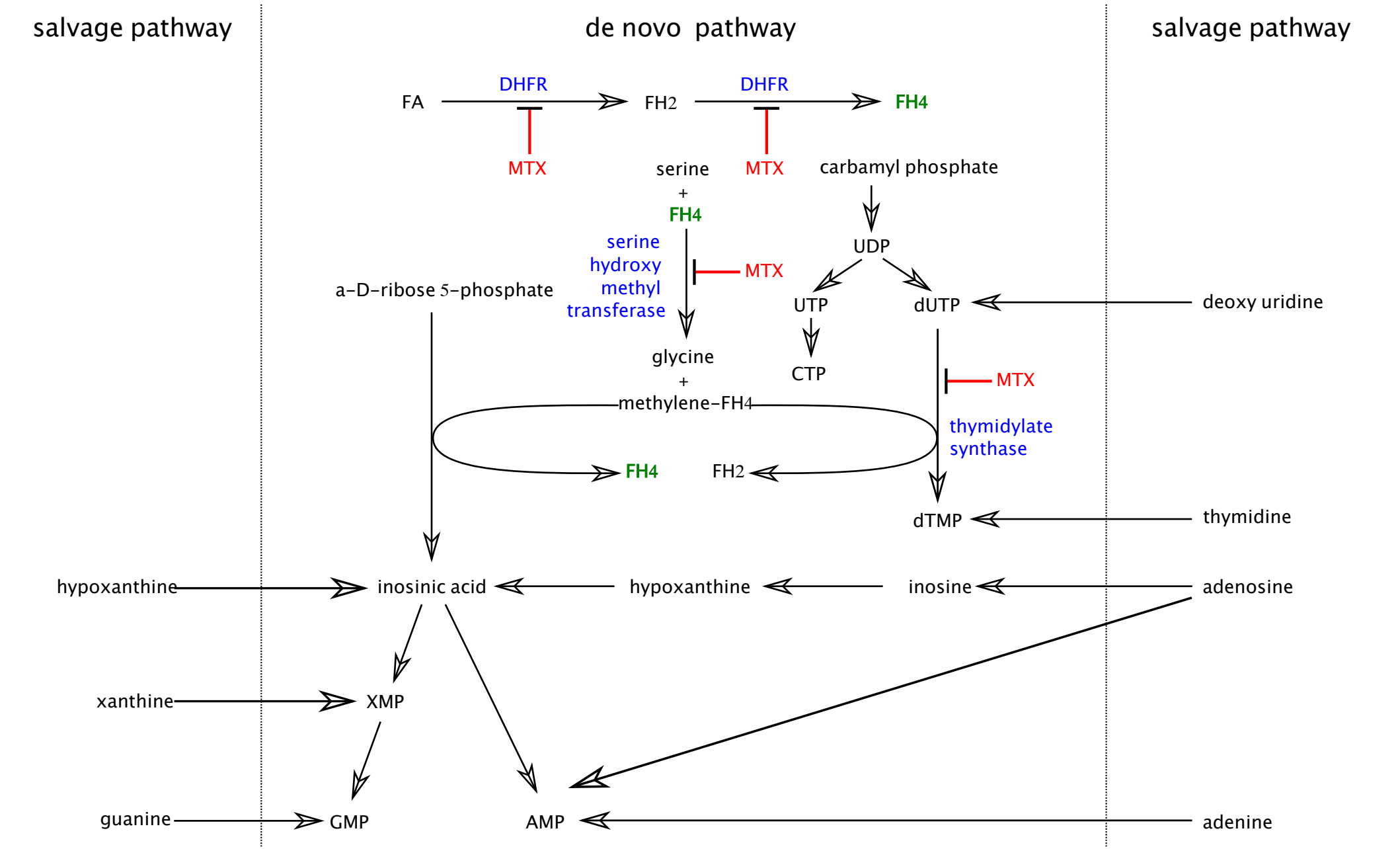
Production of recombinant proteins in mammalian ho

- **Typical product titre: ~50mg/l (1986) vs. 4.7g/l (2004)**
- **Sources of improvement:**
  a) *Cell density* is 4-5x higher due to medium & fermentation optimization (e.g. roller-bottles vs. fermenter), host cell engineering
  b) *Production phase* is 2-3x longer due to host cell engineering and fermentation optimization
  c) *Specific productivity* is 5-10x higher due to HT screening, host-cell engineering (site-specific integration, protein folding and posttranslational modification enhancements)
- Original CHO system[1,2] has no IP issues, therefore it is still frequently used!

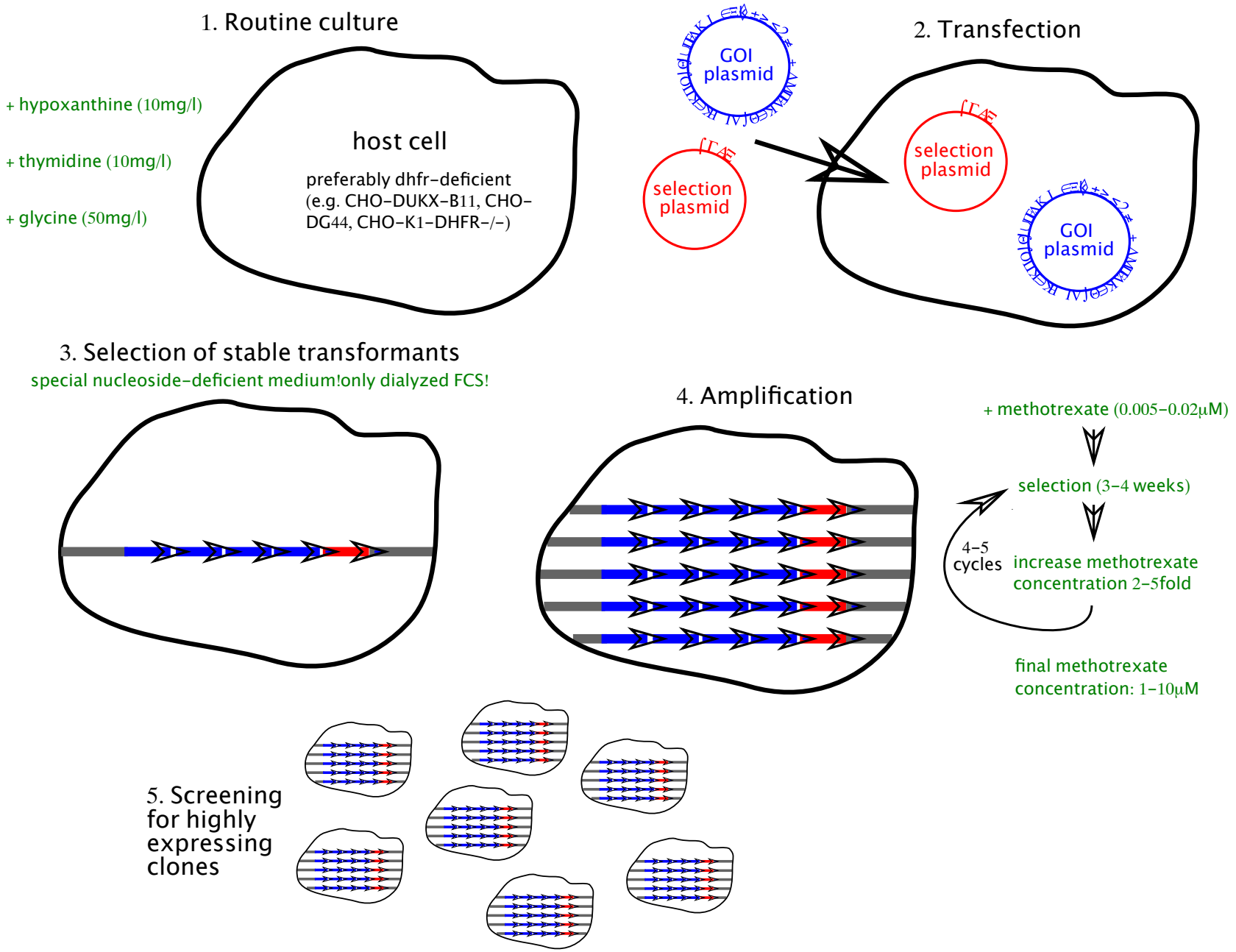
1. Kaufman, R et al. (1985) Coamplification and coexpression of human tissue-type plasminogen activator and murine dihydrofolate reductase in CHO cells. Mol Cell Biol 5, 1750-59.

2. Kingston RE et al. (2002) Amplification using CHO cell expression vectors. Curr Protoc Mol Biol Chapter 16: Unit 16.23.

DHFR (dihydrofolate reductase) metabolism



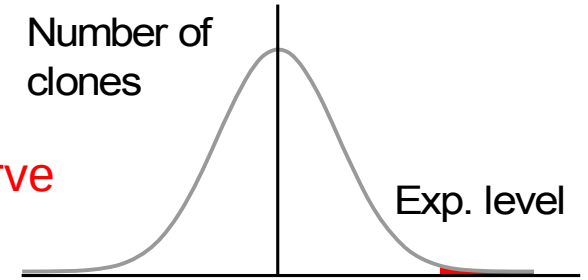
The CHO DHFR selection/amplification system



Major challenges of dhfr gene amplification in CHO

1. During selection and amplification low-expressing clones are preferentially selected (inverse relationship between protein expression levels and doubling time)
- Early and multiple rounds of single-cell expression screening
- Expression feedback (e.g. expressing the selection marker as a fusion protein with an EC domain capable of autocrine survival signalling)

2. Resistance against methotrexate (MTX) does occur due to alternative drug-resistance mechanisms, especially under rapid selection
- Early and multiple rounds of single-cell expression screening
- Slow selection



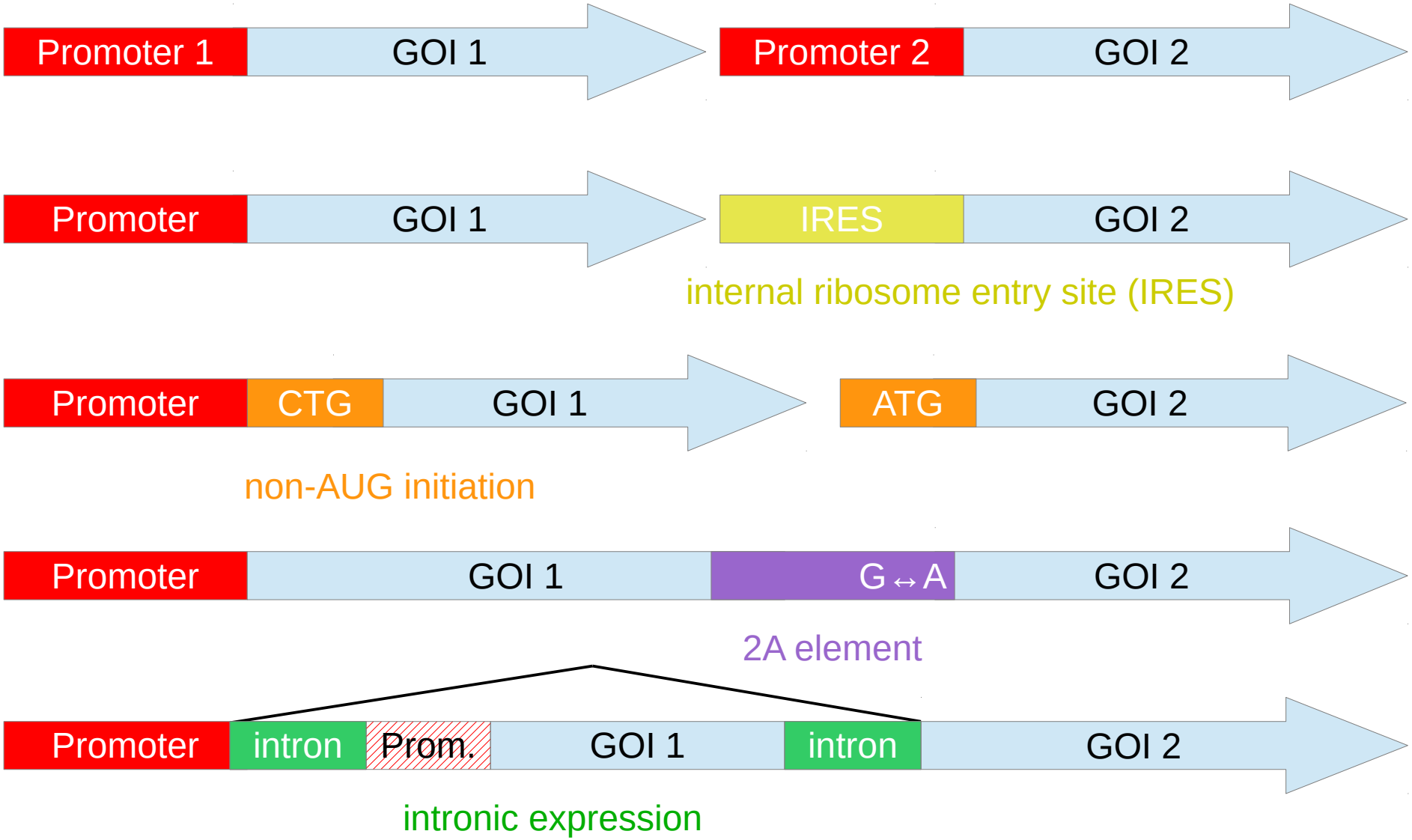
3. Finding a highly expressing clone from the tail of the Gauss curve
- Targeted insertion
- HT screening
- Equalizing to high expression levels by anti-repressing elements and CUG-translational in

4. Protein expression levels are inversely related to protein quality
- Engineering of host cells (enzymes of the glycosylation machinery and chaperones/foldas

5. Gene silencing
- Stuffer sequences, anti-silencing and anti-repressing elements

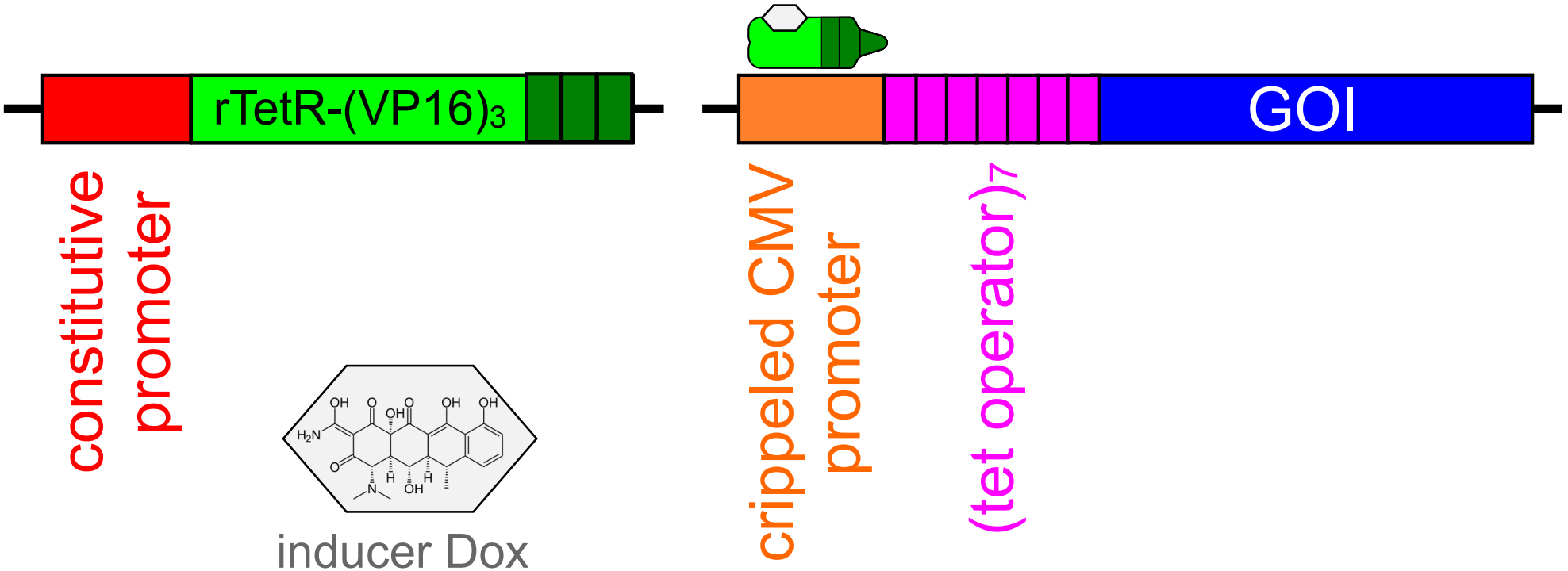
6. Resource constraints
- Up to half a year for the establishment of a highly expressing cell line

Expression of two or more genes from one plasmid



Ratio of expression between GOI1 and GOI2 can be regulated:
IRES < cap-dependent translation, CTG << ATG

Inducible expression in eukaryotic cell culture: Tet-System



inducer Dox (doxycycline)

Dox (Tet-On, Clontech; T-Rex, LifeTechnologies)
IPTG (LacSwitch, Agilent)
Edcysone (Complete Control, Agilent)
Mifepristone (GeneSwitch, LifeTechnologies)
RSL1 (RheoSwitch, NEB)
Cumate (Cumate gene switch)
Me^{2+} (methallothionine promoter)
...

Comparison of expression systems

Characteristics	***Bacteria***	***Yeast***	***Insect cells***	***Mammalian cells***
Cell cycle	very fast	fast	slow	slow
Cost of growth medium	low	low	high	high
Expression level	high	low - high	low - high	low - moderate
Secreted expression	to periplasm	yes	yes	yes
Posttranslational modifications				
Protein folding	refolding often required	refolding rarely required	mostly proper folding	proper folding
cysteine bond formation	in periplasm **(1)**	yes	yes	yes
N-linked glycosylation	no **(2)**	high mannose type	simple, no sialic acid **(3)**	complex
O-linked glycosylation	no **(2,4)**	yes	yes	yes
Phosphorylation	no **(2,4)**	yes	yes	yes
N-terminal acetylation (protein stability)	no **(5)**	yes	yes	yes
Acylation	no	yes	yes	yes
γ-carboxylation (blood clotting cascade)	no	no	no	yes

1 execpt for modified strains, **2** under development, **3** except for Sf9 mimick™, **4** with exceptions, **5** yeast-NatB-transgenic E.coli

The Cloning Challenge

- Postponed to next week!