Recombinant protein expression in bacteria

Lecture 2

Literature

Appl Microbiol Biotechnol (2006) 72:211–222
DOI 10.1007/s00253-006-0465-8

MINI-REVIEW

Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems

Kay Terpe

- Additional reading: Glick, Bernard J. (Author). Molecular Biotechnology : Principles and Applications of Recombinant DNA (4th Edition). Washington, DC, USA: ASM Press, 2010. Chapter 6

Overview of lecture

- General considerations
- Expression of proteins in *E. coli*
  - Promoters
  - Vectors
  - Plasmids systems, copy number, compatibility groups
  - Expression strains
- Alternative bacterial expression systems
  - *Bacillus subtilis*
  - *Lactococcus lactis*

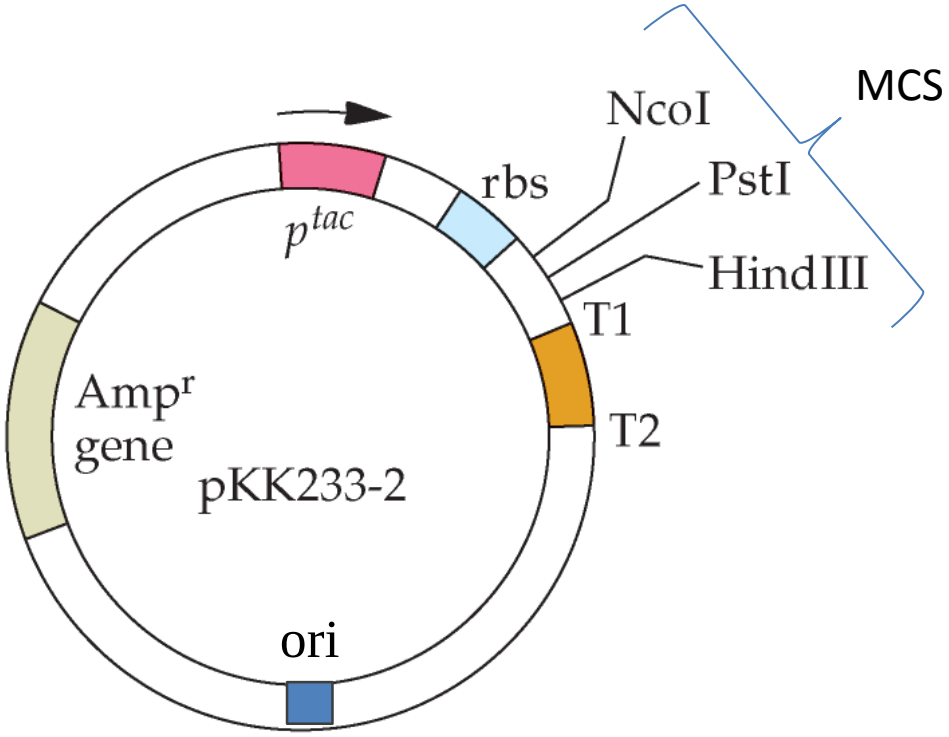
Advantages of prokaryotic expression systems

- Fast growth
- Cheap medium and equipment for growing
- Good knowledge of the host
  - Well characterized, genome known, physiology, etc.
- Tools and methods for genetic engineering available
- For therapeutic protein: no contaminations of cell cultures with human pathogens

Disadvantages of prokaryotic expression systems

- Limitation for expression of eukaryotic proteins
  - different frequencies with which the different codons appear in genes of these organisms
  - Absence of (differences in) post-translational modifications
  - S-S bonds only formed in periplasmic space
  - Protein N-glycosylation
- Might contain endotoxins (lipopolysaccharides), gram$^-$

Prokaryotic expression plasmid



- Episomal element
- Prokaryotic promoter
- RBS: ribosome binding site
- MCS: multiple cloning site
- transcription termination site (T1 and T2)
- Prokaryotic selectable marker (e.g. ampicillin Amp^R)
- Origin of replication (ori)
- Gene with start codon

- Where does transcription & translation start?

Considerations for choice of promoter

- Controllability of expression
  - Constitutive vs. inducible
  - Leakiness
  - Range of induction levels
- Method for induction
  - Inducing agent (IPTG, sugars, antibiotics)
    - Price of inducers
    - Regulatory objections
  - Change of cultivation condition (temperature, oxygen level, pH, starvation)
    - Are these changes compatible with growth or protein expression?

Origin of replication

- Origin of replication is often host-specific
- The number of plasmids per cell is determined by the origin of replication
- Relaxed plasmids produce many copies of themselves per host cell, while stringent plasmids produce only a few copies per cell
- Broad host range origins of replication available, which enable the use of a plasmid in different organisms, e.g. RK2 in most gram$^-$ bacteria

Replicon and copy number

- Maintaining plasmids is a cellular burden!
  - *E. coli* genome: 4,639,221 base pairs
  - Average plasmid size: 10’000 to 5’000 base pairs
- Cells harboring plasmids grow slower than plasmid free control cells, even if plasmid does not express a protein

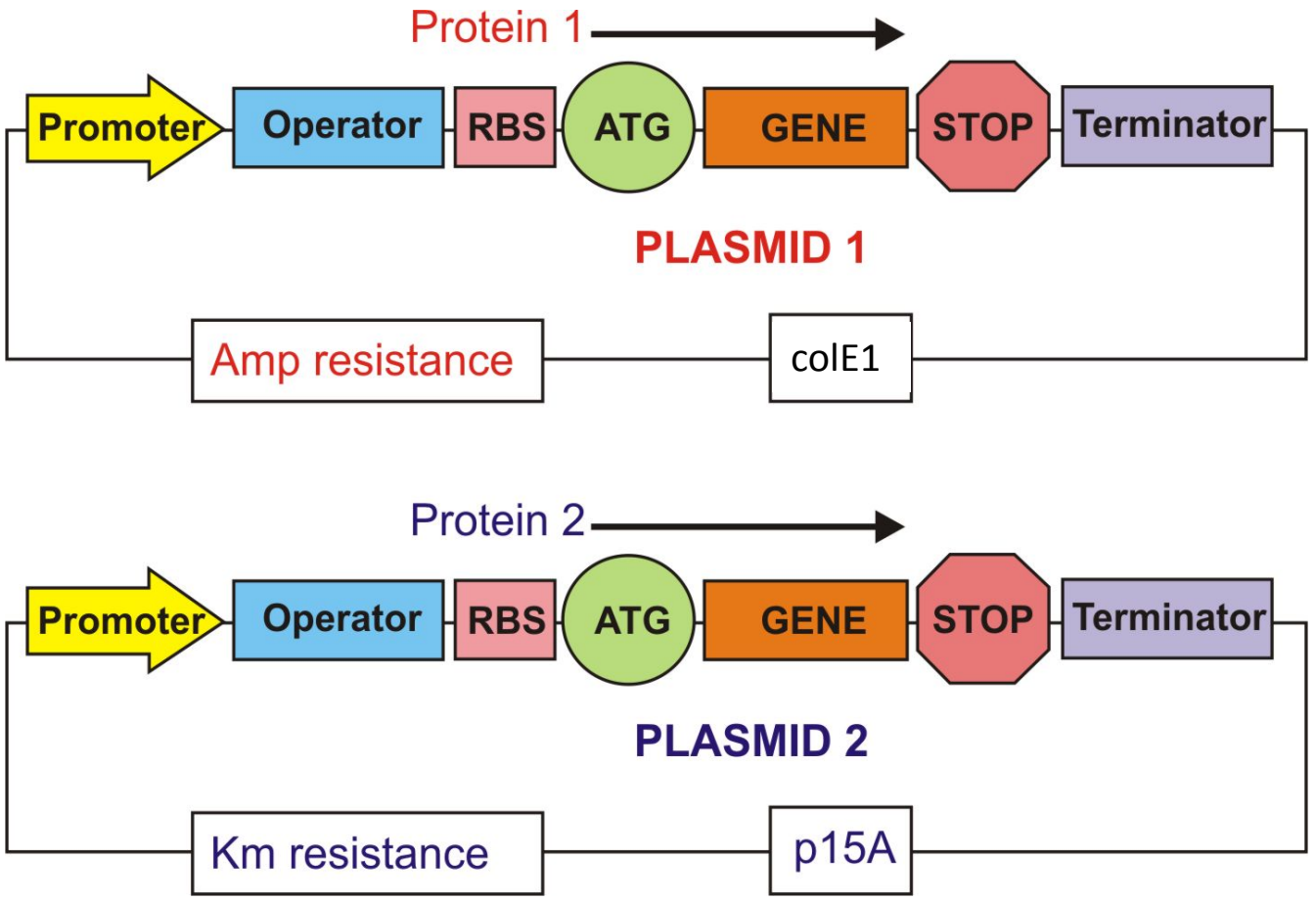
Plasmid	Replicon	Copy Number
pBR322	pMB1	15 - 20
pUC19	modified pMB1	500 - 700
pMOB45	pKN402	15 - 120
pACYC	p15A	18 - 22
pSC101	pSC101	~5
colE1	colE1	15 - 20

Replicon and incompatibility group

- In addition to the number of plasmids per cell, the mechanism for plasmid replication needs to be considered
  - Plasmids belong to different incompatibility groups
- Two plasmids from the same incompatibility group cannot be stably maintained in a single cell
  - e.g. plasmids with colE1 and pMB1 are incompatible
- Incompatibility is based on the mechanism for controlling plasmid replication

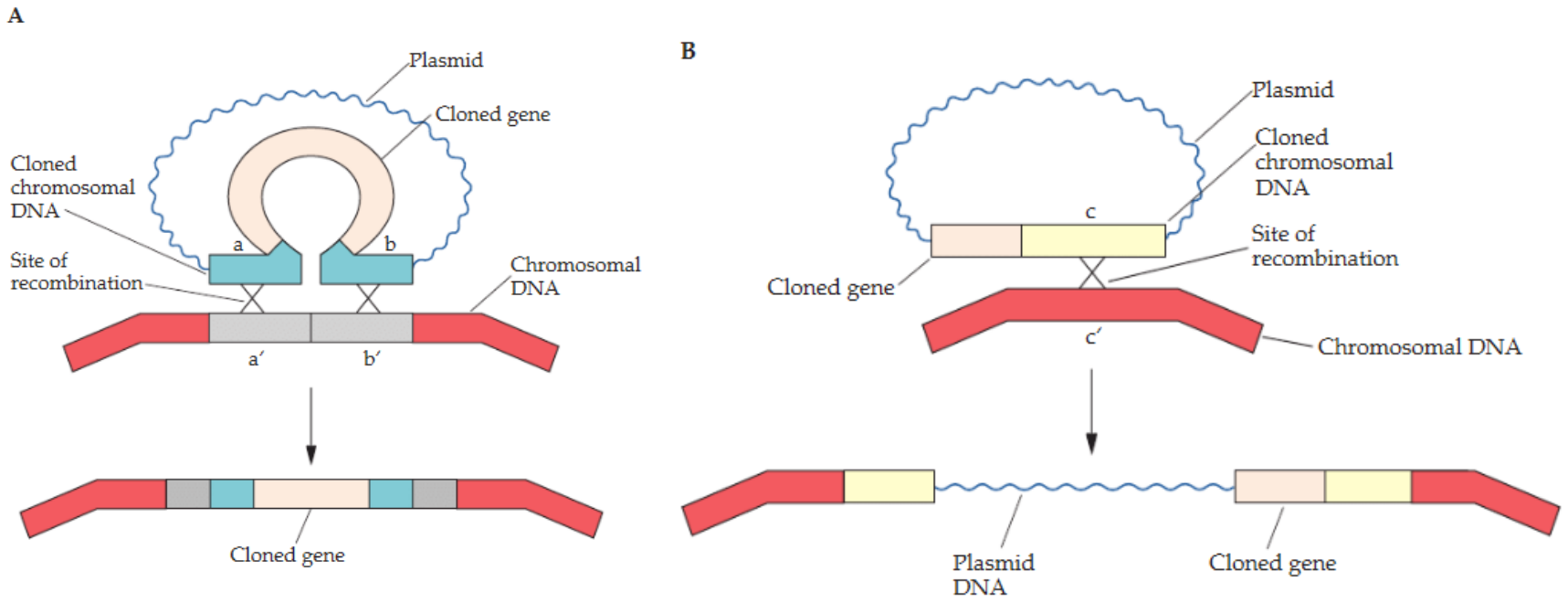
Incompatibility groupings	**Negative control elements**	**Mechanism of action**
colE1, pMB1	RNA I (different from RNAi)	controls processing of pre-RNAII into primer
IncFII, pT181	RNA	controls synthesis of RepA protein
P1, F, R6K, pSC101, p15A	interons	sequesters RepA protein

Co-expression from two plasmids



Two different selection markers and origin of replication required

Integration of DNA into chromosome

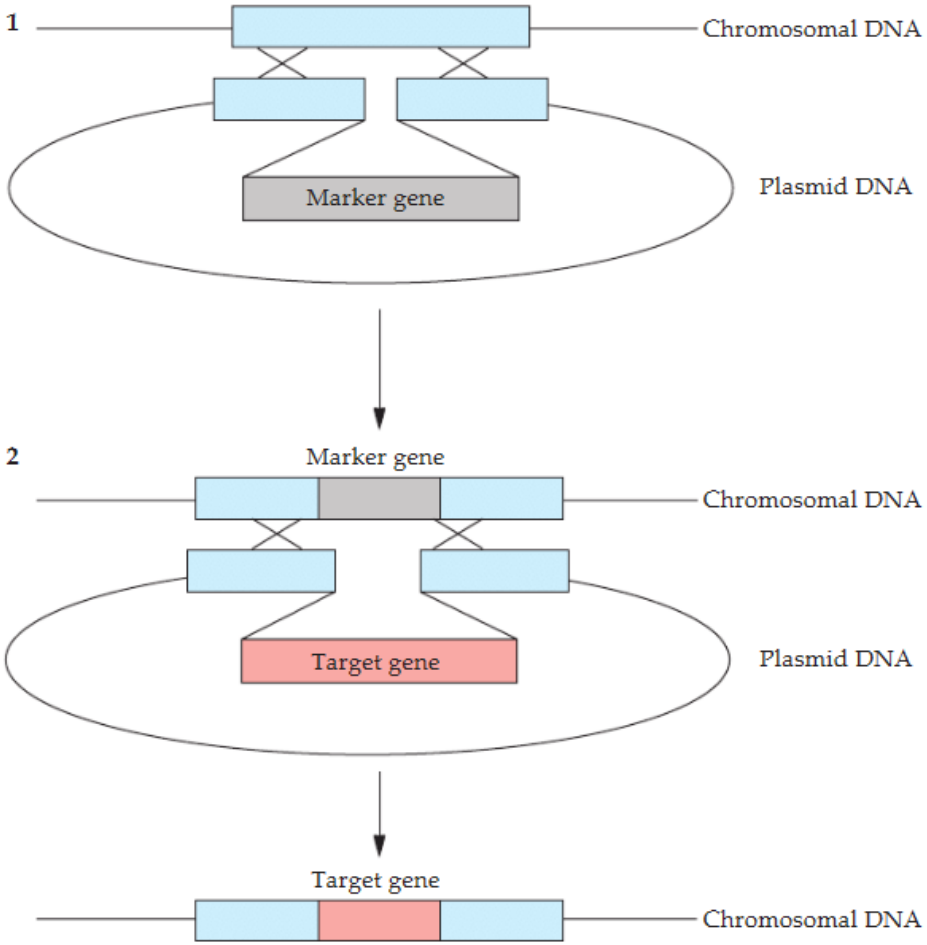


- Integration of a cloned gene into the chromosome. A) via double cross-over B) via single cross-over event
- Besides the cloned gene, most likely also selection marker will be introduced!
- Nor origin of replication present on integrative plasmid!

Strategy for generation of a marker-free strain

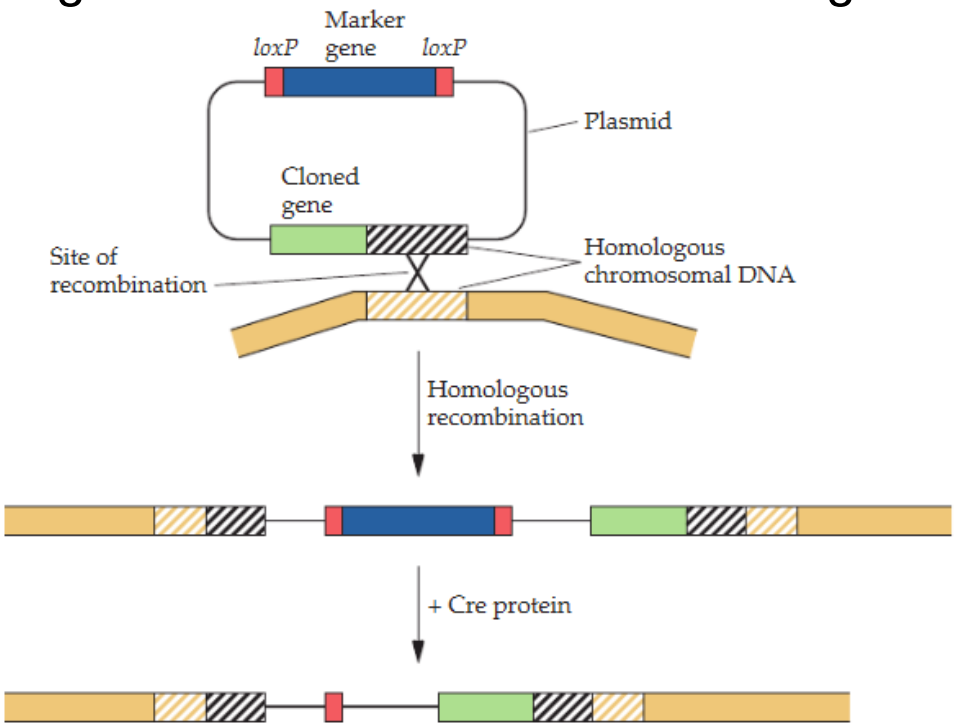
Insertion of marker into predetermined site on chromosome

Selectable marker gene is replaced by target gene



Cre-lox system for removing marker

- The system consists of the Cre recombinase, that recombines a pair of short target sequences called the *LoxP* sequences.
- The Cre enzyme and the *LoxP* sequence are derived from bacteriophage P1
- The Cre enzyme is expressed in the host organism after integration of recombinant DNA into genome



1. step: Integration of recombinant DNA into genome

2. step: expression of Cre enzyme
3. step: removal of DNA between loxP sites

Escherichia coli as expression system

- Good knowledge of the host (genome, physiology)
- Tools for genetic engineering established
- Fast growth
- Cheap medium and equipment for growing
- GRAS (generally regarded as safe) status (depends on strain used)
- Up to 50% of total cellular protein content can be the recombinant protein

Disadvantages of *E. coli*

- Need to eliminate endotoxin from products
- Cells must be lysed to get product (inner and outer membrane)
- Considered not to be a good “secretory system”
  - Secretion into periplasmic space
  - Periplasmic space measures only 8 to 16% of total cell volume
- Limitation for expression of eukaryotic proteins due to:
  - different frequencies with which the different codons appear in genes of these organisms
  - Absence of (differences in) post-translational modifications
  - S-S bonds only formed in periplasmic space
  - Protein N-glycosylation

Available promoters

- Endogenous host's promoters
  - 2500 in the entire genome of *E. coli* K12 strain
- Most frequently used: P_{lac} / P_{tac} / P_{trc}, P_{BAD}, rhaP_{BAD}
  - Chemically inducible gene expression

- Promoters from phages
  - T7, T3, SP6, T5, PL
  - Highly efficient and specific expression

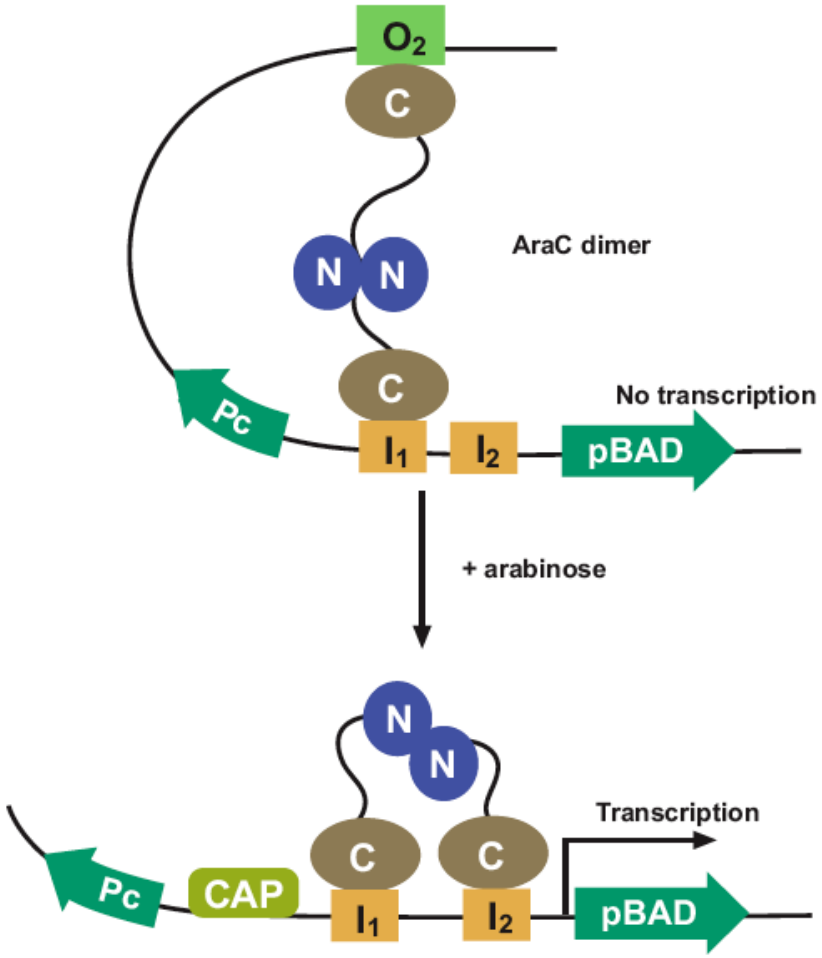
Table 4 Some *E. coli* promoter systems that are in use for heterologous protein production and their characteristics

Expression system based on	Induction (range of inductor)	Level of expression	Key features
lac promoter	Addition of IPTG 0.2 mM (0.05–2.0 mM)	Low level up to middle	Weak, regulated suitable for gene products at very low intracellular level Comparatively expensive induction
trc and *tac* promoter	Addition of IPTG 0.2 mM (0.05–2.0 mM)	Moderately high	High-level, but lower than T7 system Regulated expression still possible Comparatively expensive induction High basal level
T7 RNA polymerase	Addition of IPTG 0.2 mM (0.05–2.0 mM)	Very high	Utilizes T7 RNA polymerase High-level inducible over expression T7*lac* system for tight control of induction needed for more toxic clones Relative expensive induction Basal level depends on used strain (pLys)
Phage promoter p_L	Shifting the temperature from 30 to 42 °C (45 °C)	Moderately high	Temperature-sensitive host required Less likelihood of "leaky" uninduced expression Basal level, high basal level by temperatures below 30 °C No inducer
tetA promoter/operator	Anhydrotetracycline 200μg/l	Variable from middle to high level	Tight regulation Independent on metabolic state Independent on *E. coli* strain Relative inexpensive inducer Low basal level
araBAD promoter (P_{BAD})	Addition of L-arabinose 0.2 % (0.001–1.0 %)	Variable from low to high level	Can fine-tune expression levels in a dose-dependent manner Tight regulation possible Low basal level Inexpensive inducer
rhaP$_{BAD}$ promoter	L-rhamnose 0.2 %	Variable from low to high level	Tight regulation Low basal activity Relative expensive inducer

Alternative promoters for expression in *E. coli*

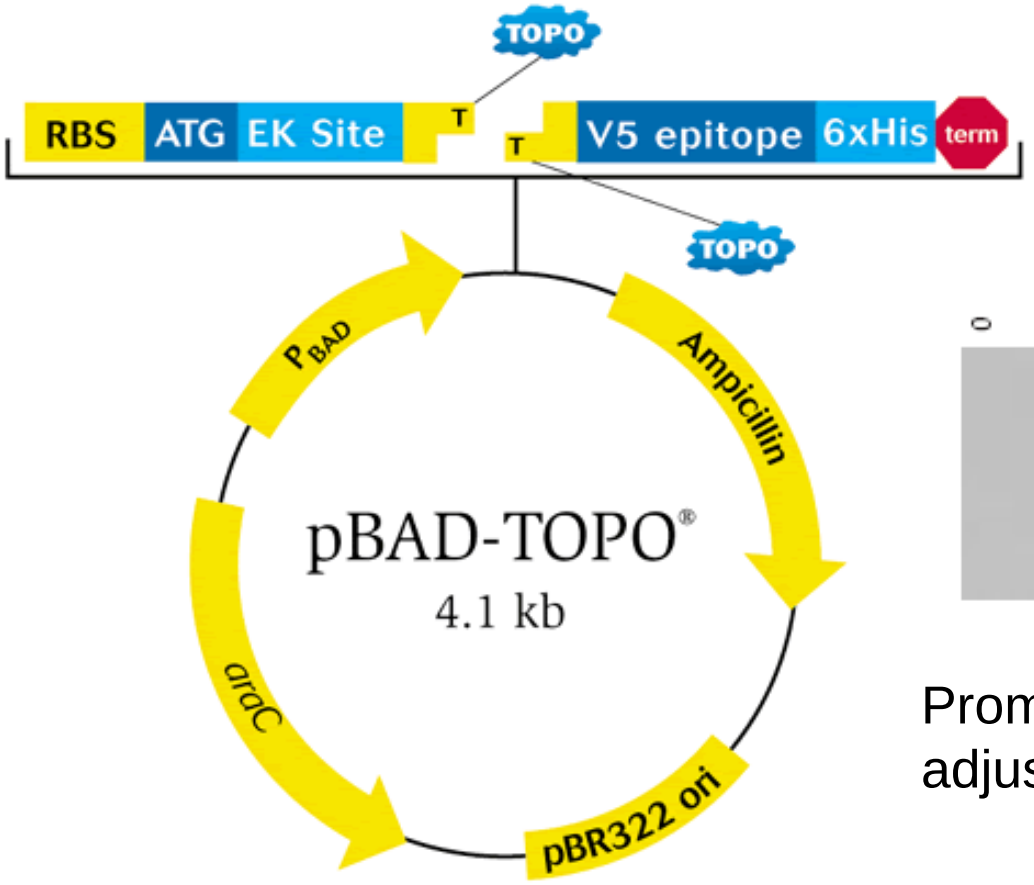
Promoter	Induction	Level of Expression	Additional Information
P_{VHb}	Oxygen limitation	up to 20% of total protein	• dissolved oxygen levels of less than 2% air saturation
P_{cspA}	Thermal cold shock (shift to below 20°C)		• The cspA core promoter is only weakly induced by temperature downshift. • A 159 nucleotide long untranslated region at the 5' end of cspA driven transcripts makes them highly unstable at 37°C and significantly increases their stability at low temps. • cspA system becomes repressed 1-2 hours after temperature downshift.

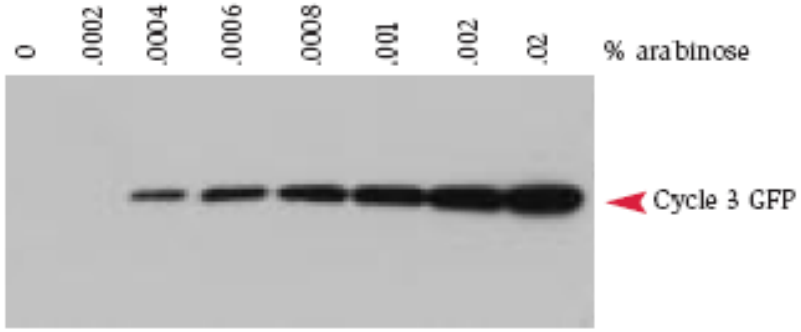
Arabinose inducible P_{BAD} promoter



- Regulated by the transcriptional regulator *AraC*
  - *AraC* forms a complex with L-arabinose
  - AraC has varying affinity for O_2, I_1 and I_2
- <u>absence of arabinose</u>
  - AraC dimer contacts the O_2 and I_1 half site of the operon, forming a 210 bp DNA loop.
- <u>presence of arabinose</u>
  - Arabinose binds to AraC. AraC releases the O_2 site and binds the I_2 site, which is adjacent to the I_1 site. This releases the DNA loop and allows transcription to begin.
  - The cAMP activator protein (CAP)-cAMP complex binds to the DNA and stimulates binding of AraC to I_1 and I_2.

P_{BAD} vector





Promoter titratable allowing exact adjustment of expression levels

Regulation of gene expression by the p^L promoter of bacteriophage λ

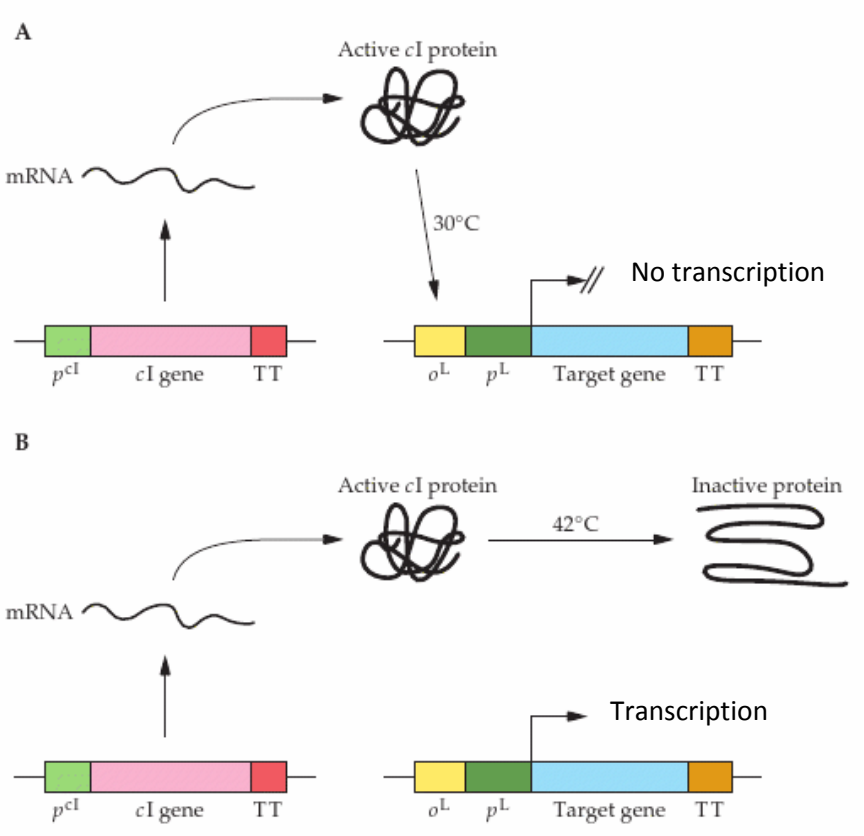
- The p^L promoter is controlled by the cI repressor protein of bacteriophage λ
- A temperature-sensitive mutant of the cI repressor is used to regulate p^L-directed transcription.
- the cI repressor is stable at 28 to 30°C and prevents transcription from the p^L promoter.
- When the temperature is shifted to 42°C, the thermosensitive cI repressor is inactivated and transcription can proceed.

Regulation of gene expression by the pL promoter

(A) At 30°C, the cI repressor, which is synthesized constitutively under the control of its own promoter (p^{cI}), binds to the operator region (o^{L}) of the p^{L} promoter, preventing the target gene from being transcribed

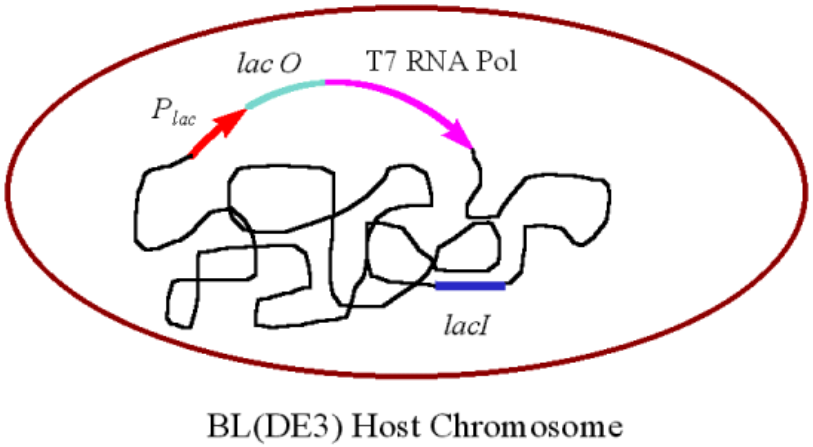
(B) At 42°C, the temperature sensitive cI repressor is synthesized and inactivated so that it no longer interferes with transcription of the target protein

Plasmids for expression of repressor and of the target gene are required

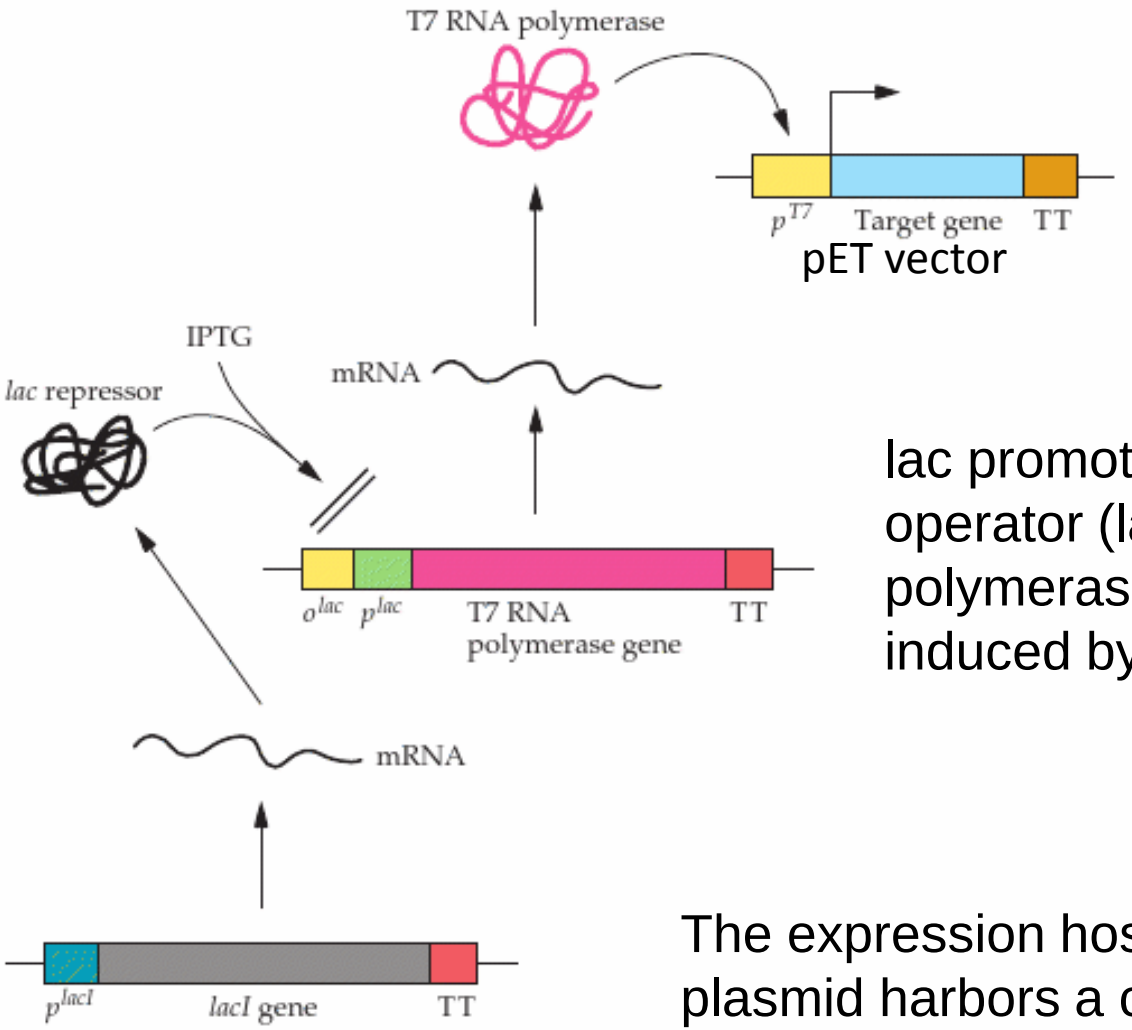


The pET expression system

- The gene of interest is under control of a T7 bacteriophage promoter
- Promoter is only recognized by T7 RNA polymerase
  - Prokaryotic cells do not produce T7 RNA polymerase
- Host cell genetically engineered to incorporate the gene for T7 RNA polymerase
  - T7 RNA polymerase under control of lac promoter (P_{lac}) and lac operator (lac O)
  - Lac repressor (lacI)
- A typical host strain with pET system is *E. coli* strain BL(DE3).



The pET expression system



p^{T7} promoter specific for T7 polymerase

lac promoter (P_{lac}) and lac operator (lac O) control T7 RNA polymerase gene, expression induced by addition of IPTG

The expression host and each pET plasmid harbors a copy of lacI gene

E. coli host strain used

- Choosing the proper *E. coli* host strain is obligatory for substantial expression of recombinant proteins
  - It should be protease deficient
  - The most commonly used protease deficient strain is *E. coli* BL21
- Engineered strains
  - Expressing rare codons
  - Enabling formation of S-S bonds in cytoplasm

E. coli strains used for protein expression

Table 1 Some *E. coli* strains most frequently used for heterologous protein production and their key features

E. coli strain	Derivation	Key features
AD494	K-12	*trxB* mutant; facilitates cytoplasmic disulfide bond formation
BL21	B834	Deficient in *lon* and *ompT* proteases
BL21 trxB	BL21	*trxB* mutant; facilitates cytoplasmic disulfide bond formation; deficient in *lon* and *ompT* proteases
BL21 CodonPlus-RIL	BL21	Enhances the expression of eukaryotic proteins that contain codons rarely used in *E. coli*: AGG, AGA, AUA, CUA; deficient in *lon* and *ompT* proteases.
BL21 CodonPlus-RP	BL21	Enhances the expression of eukaryotic proteins that contain codons rarely used in *E. coli*: AGG, AGA, CCC; deficient in *lon* and *ompT* proteases.
BLR	BL21	*recA* mutant; stabilizes tandem repeats; deficient in *lon* and *ompT* proteases
B834	B strain	Met auxotroph; ^{35}S-met labeling
C41	BL21	Mutant designed for expression of membrane proteins
C43	BL21	Double mutant designed for expression of membrane proteins
HMS174	K-12	*recA* mutant; Rif resistance
JM 83	K-12	Usable for secretion of recombinant proteins into the periplasm
Origami	K-12	*trxB/gor* mutant; greatly facilitates cytoplasmic disulfide bond formation
Origami B	BL21	*trxB/gor* mutant; greatly facilitates cytoplasmic disulfide bond formation; deficient in *lon* and *ompT* proteases
Rosetta	BL21	Enhances the expression of eukaryotic proteins that contain codons rarely used in *E. coli*: AUA, AGG, AGA, CGG, CUA, CCC, and GGA; deficient in *lon* and *ompT* proteases
Rosetta-gami	BL21	Enhances the expression of eukaryotic proteins that contain codons rarely used in *E. coli*: AUA, AGG, AGA, CGG, CUA, CCC, and GGA; deficient in *lon* and *ompT* proteases; *trxB/gor* mutant; greatly facilitates cytoplasmic disulfide bond formation

Most strains are also available as DE3 and DE3 pLysS strains. Strains are commercially available from different manufacturers

Potential problems of protein expression in *E. coli*

Table 2a Some problems of heterologous protein production in *E. coli* and possible solutions

Symptom	Possible problem	A collection of solutions
Cell death or no colonies	Toxic protein, high basal expression	More stringent control over basal expression Tightly controlled promoter system Weaker promoter Lowering temperature Lowering inducer concentration
Insoluble disulfide protein (inclusion bodies)	Reduction of disulfide bonds	Minimize reduction in cytoplasm Accumulation in the periplasm
Insoluble protein (inclusion bodies)	Too much expression	Attenuate expression by: weaker promoter, lowering temperature, lowering inducer concentration, decrease plasmid copy number, fusion of a hydrophilic affinity tag
No activity	Misfolded protein, affinity tag can decrease activity	Minimize reduction in cytoplasm Accumulation in the periplasm Attenuate expression Change affinity tag
No protein, truncated protein	*E. coli* codon usage (codon bias)	Supply rare tRNAs Stronger promoter Increase plasmid copy number Lower temperature Tightly controlled promoter system

Nevertheless, another bacterial host than *E. coli* could also solve the problem

Reasons for choosing alternative expression hosts

- Rarely used codons or endotoxins could be a reason to change from an *E*. *coli* system to another host
- Principally, all bacteria could be used for heterologous protein production
- The genomics era offers new information about the bacteria hosts that are rarely used
- Another attractive reason for pharmaceutical industry to work with another host or new promoter system could be the patent situation

Bacilli as expression strains

Table 5 Some pharmaceutically and industrially relevant proteins that were successfully expressed in different Bacilli strains

Recombinant protein	Bacillus strain	Yield (mg/l)	Reference
α-amylase (*Bacillus amyloliquefaciens*)	*B. subtilis*	1,000–3,000	Palva (1982)
α-amylase (*Bacillus stearothermophilus*)	*B. brevis*	3,000	Udaka and Yamagata (1993)
α-amylase (human)	*B. brevis*	60	Konishi et al. (1990)
Cellulase	*B. brevis*	100	Kashima and Udaka (2004)
Cholera toxin B	*B. brevis*	1400	Ichikawa et al. (1993)
Dextransucrase (*Leuconostoc mesenteroides*)	*B. megaterium*	n.d. (362 U/g)	Malten et al. (2005)
Gelatin	*B. brevis*	500	Kajino et al. (2000)
Epidermal growth factor (human)	*B. brevis*	240	Yamagata et al. (1989)
Epidermal growth factor (human)	*B. subtilis*	7	Lam et al. (1998)
Epidermal growth factor (mouse)	*B. brevis*	50	Wang et al. (1993)
Mouse/human chimeric Fab′	*B. brevis*	100	Inoue et al. (1997)
Interferon-α2 (human)	*B. subtilis*	0.5–1.0	Palva et al. (1983)
Interleukin-2 (human)	*B. brevis*	120	Takimura et al. (1997)
Interleukin-6 (human)	*B. brevis*	200	Shiga et al. (2000)
Lipase A	*B. subtilis*	600	Lesuisse et al. (1993)
Penicillin G acylase	*B. subtilis*	n.d.	Yang et al. (2001)
Pepsinogen (swine)	*B. brevis*	11	Udaka and Yamagata (1993)
PHA depolymerase A (*Paucimonas lemoignei*)	*B. subtilis*	1.9	Braaz et al. (2002)
Proinsulin	*B. subtilis*	1000	Olmos-Soto and Contreras-Flores (2003)
Protein disulfide isomerase	*B. brevis*	1100	Kajino et al. (1999)
ScFv	*B. subtilis*	10–15	Wu et al. (2002)
Staphylokinase	*B. subtilis*	337	Ye et al. (1999)
Streptavidin	*B. subtilis*	35–50	Wu and Wong (2002)
Thioredoxin (*Alicyclobacillus acidocaldarius*)	*B. subtilis*	500	Anna et al. (2003)
Toxin A (*Clostridium difficile*)	*B. megaterium*	n.d.	Burger et al. (2003)

B. subtilis expression system

- Gram^{+} bacteria are well known for their contributions to agricultural, medical and food biotechnology and for the production of recombinant proteins.
- *Bacilli* strains have no lipopolysaccharides (endotoxin-free)
- *Bacilli* strains have a naturally high secretion capacity, and export proteins directly into the extracellular medium
- little is known about disulfide bond formation and isomerization

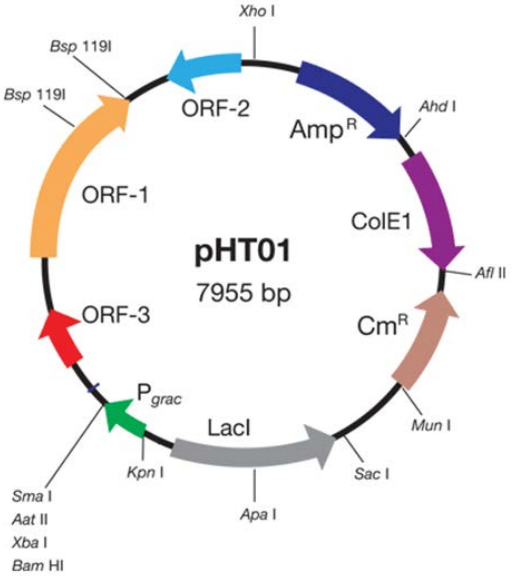
B. subtilis expression system

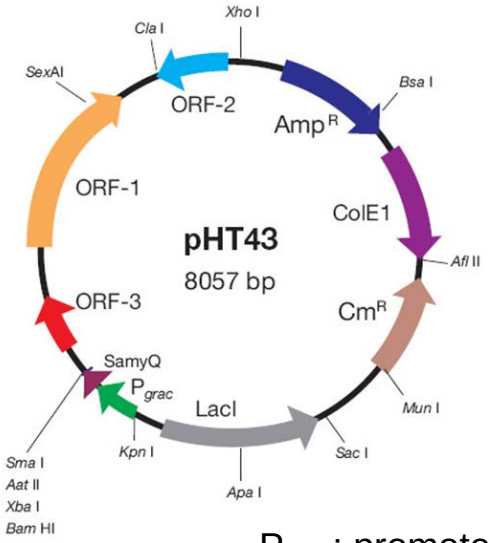
- *B. subtilis* has been developed as an attractive host because of several reasons:
  - non-pathogenic and is considered as a GRAS organism (generally regarded as safe)
  - It has no significant bias in codon usage
  - capable of secreting functional extracellular proteins directly into the culture medium (at present, about 60% of the commercially available enzymes are produced by Bacillus species)
  - large body of information concerning transcription, translation, protein folding and secretion mechanisms, genetic manipulation and large-scale fermentation has been acquired.

Limitations B. subtilis system

- production of a number of extracellular proteases which recognize and degrade heterologous proteins
  - First obstacle has been largely solved by the construction of protease-deficient strains.
- availability of stable plasmids
  - Development of plasmids using the theta-mode of replication such as those derived from the natural plasmids pAMβ1 and pBS72
- transformation is more difficult
  - Polyethylene glycol-mediated protoplasting is necessary for efficient transformation

Intracellular and secretory expression





P_{grac}: promoter
SamyQ: secretion signal sequence
LacI: Lac repressor

- pHT43 allows high-level expression of recombinant proteins within the cytoplasm
- pHT43 directs the recombinant proteins into the medium
- Both promoters are based on the strong σ^A-dependent regulatory sequences preceding the groE operon of *B. subtilis* to which the *lac* operator has been added to render it IPTG-inducible

Mo Bi Tec
MOLECULAR BIOLOGISCHE TECHNOLOGIE

Bacillus subtilis host strains

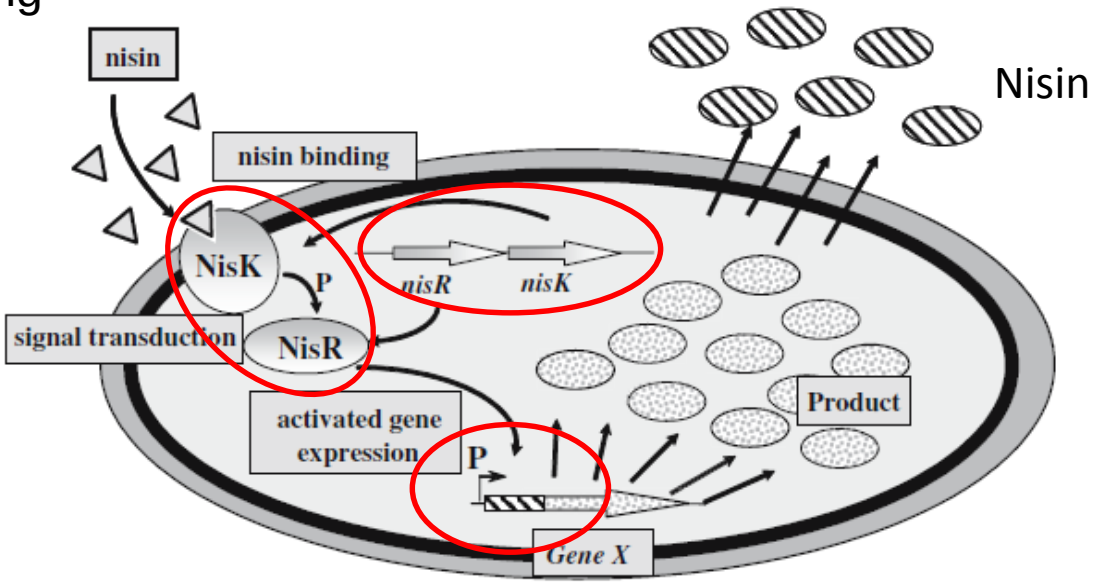
- *Bacillus subtilis* strains suitable as hosts for gene expression are available:
- For intracellular expression:
  - 1012 wild type: leuA8 metB5 trpC2 hsdRM1 (commonly used)
  - 168 Marburg: trpC2 (Trp^-)
- for secretion -> protease deficient strain
  - WB800N: nprE aprE epr bpr mpr::ble nprB::bsr vpr wprA::hyg cm::neo; NeoR

Lactococcus lactis

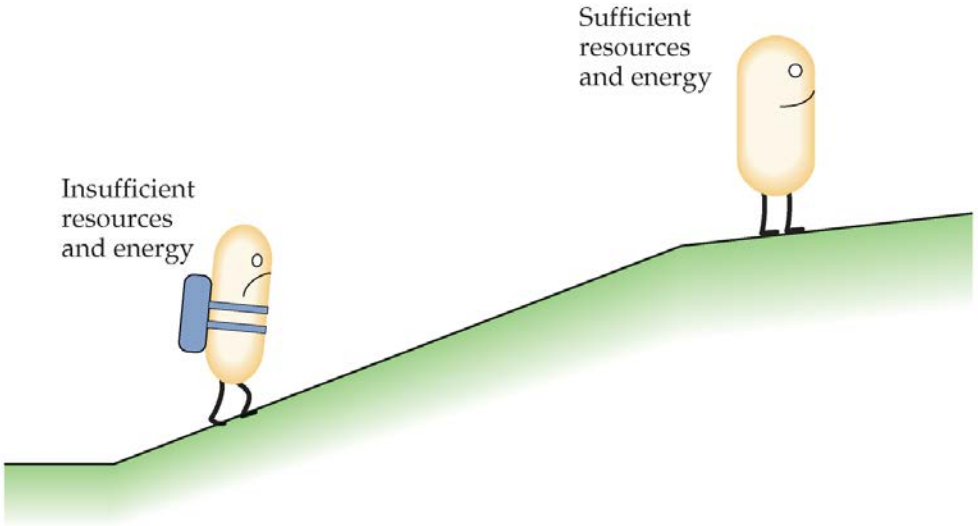
- *L*. *lactis* is a homofermentative bacterium with primary function to rapidly produce lactic acid from lactose
- *L*. *lactis* system is fully food grade
- Functional characteristics that have extensively been studied in lactococci include the carbon metabolism, the extracellular and intracellular proteolytic system
- use of lactococci in several fields of biotechnology
  - the expression of bacterial and viral antigens for vaccination via mucosal immunization
  - the use as a cell factory for the production of specific compounds and the pilot production of pharmaceutical products.

The nisin controlled gene expression system in *L. lactis*

- Protein expression is based on a **two-component regulatory system** comprising a sensor (NisK) and a regulator (NisR)
- The signal is passed from NisK to NisR via protein phosphorylation cascade (P)
- Phosphorylated NisR is activating the P_{nisA} promoter and initiates transcription of gene of interest
- In case protein of interest and nisin is coexpressed, the system is self-amplifying



Expressing a foreign protein generates a metabolic load on host bacteria



A high level of continuous expression of a cloned gene is often detrimental to the host cell because it creates an energy drain, thereby impairing essential host cell functions

All or a portion of the plasmid carrying a continuously (constitutively) expressed cloned gene may be lost after several division cycles, since cells without a plasmid grow faster and eventually take over the culture.