

Expressão em *E. coli*

Biotechnol. J. (2011) 6, 715-730.

Biotechnology Advances (2012) 30(5):1102-7.

Profa. Ana Paula - 2015

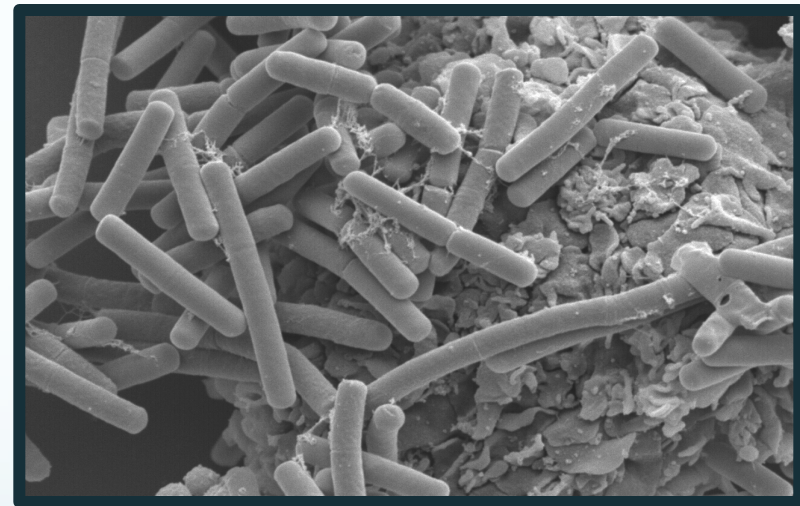
Escherichia coli

É o organismo mais comumente utilizado p/ expressão heteróloga.

Como escolher o melhor vetor? A linhagem?

- o que é importante?

Como otimizar a expressão?



Expressão em *E. coli*

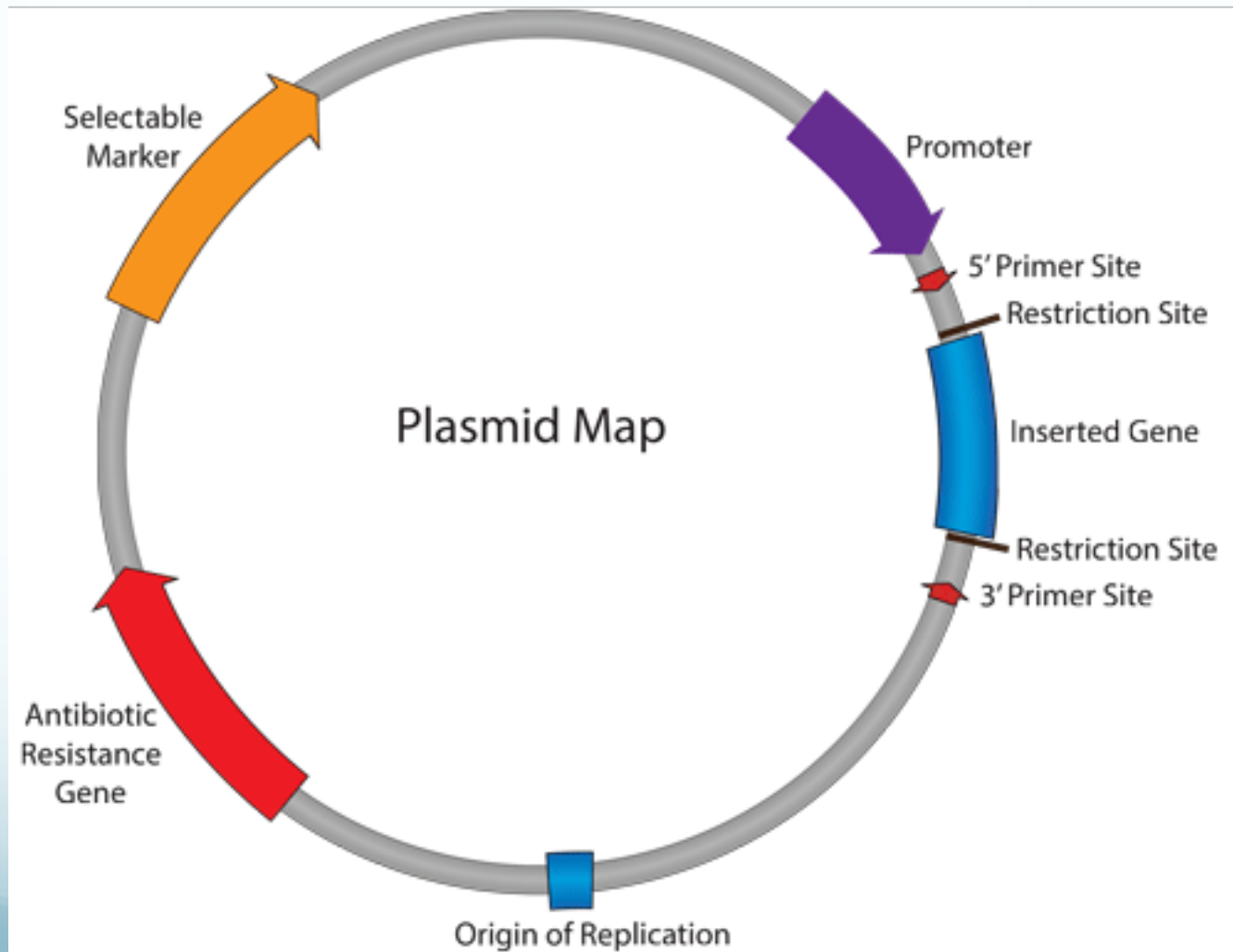
Escolhendo um vetor...

- Informações sobre a proteína alvo são essenciais
 - pI; localização subcelular, pontes S-S...

Para expressão heteróloga, os maiores “gargalos” em *E. coli* ainda são:

- Incapacidade de fazer modificações pós-traducionais;
- Capacidade limitada para catalisar pontes S-S

Os vetores de expressão



Marcas de seleção

- Mais comuns:

- **Antibióticos:** ampicilina, canamicina, cloranfenicol e tetraciclina.

Ampicilina: susceptível à degradação enzimática e à hidrólise em função do pH.

- **Sistemas de plasmídeos antibiótico-free**

- *Reduz custos e riscos de “espalhar” a resistência*
 - Deleção de genes essenciais, que são colocados no plasmídeo (ex. Fator de iniciação da tradução)

Promotores

- Ideais:
 - devem conferir altos níveis de expressão,
 - Ser induzíveis (simples e de baixo custo) e
 - Ter baixos níveis de atividade transcricional basal.

Promotores mais usados na expressão heteróloga em *E. coli*:

- *lac* (lactose; Yanisch-Perron *et al.*, 1985)
 - *trp* (triptofano, Goeddel *et al.*, 1980)
 - híbridos derivados: *tac*, *trc* (*trp/lac*)
 - araBAD (indução por L-arabinose)

Outros sistemas de expressão incluem:

promotores dos fagos: λ P_L e P_R (Elvin *et al.*, 1990);
e T7: PT7 (Tabor e Richardson, 1985).

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 inducer)	Level of expression	Key features	Original reference
<i>lac</i> 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	Gronenborn (1976)
<i>trc</i> and <i>tac</i> 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	Brosius et al. (1985)
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 <i>lac</i> system for tight control of induction needed for more toxic clones Relative expensive induction Basal level depends on used strain (pLys)	Studier and Moffatt (1986)
Phage promoter <i>p_L</i>	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	Elvin et al. (1990)
<i>tetA</i> promoter/operator	Anhydrotetracycline 200 µg/l	Variable from middle to high level	Tight regulation Independent on metabolic state Independent on <i>E. coli</i> strain Relative inexpensive inducer Low basal level	Skerra (1994)
<i>araBAD</i> promoter (<i>P_{BAD}</i>)	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	Guzman et al. (1995)
<i>rhaP_{BAD}</i> promoter	L-rhamnose 0.2 %	Variable from low to high level	Tight regulation Low basal activity Relative expensive inducer	Haldimann et al. (1998)

Elementos de um vetor de expressão procarioto

291

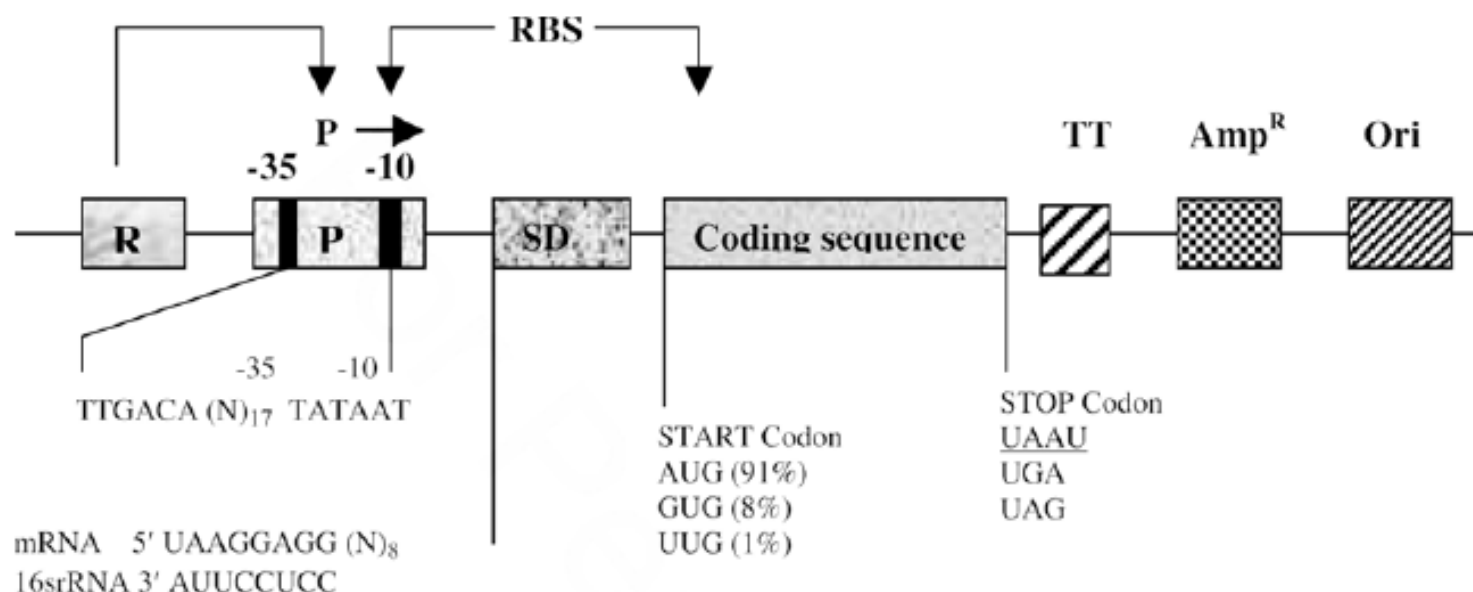


Fig. 1 Schematic presentation of the salient features and sequence elements of a prokaryotic expression vector. Shown as an example is the hybrid *tac* promoter (*P*) consisting of the 235 and 210 sequences, which are separated by a 17-base spacer. The arrow indicates the direction of transcription. The ribosome-binding site (*RBS*) consists of the Shine–Dalgarno (*SD*) sequence, followed by an A+T-rich translational spacer that has an optimal length of approximately eight bases. The *SD* sequence interacts with the 3' end of the 16S rRNA during translational initiation, as shown. The three start codons are shown, along with the frequency of their usage in *Escher-*

ichia coli. Among the three stop codons, UAA followed by U is the most efficient translational termination sequence in *E. coli*. The repressor is encoded by a regulatory gene (*R*), which may be present on the vector itself or may be integrated in the host chromosome, and it modulates the activity of the promoter. The transcription terminator (*TT*) serves to stabilize the mRNA and the vector. In addition, an antibiotic resistance gene, e.g., for ampicillin, facilitates phenotypic selection of the vector, and the origin of replication (*Ori*) determines the vector copy number. The various features are not drawn to scale

Regulação da tradução

- A iniciação da tradução requer um RBS, incluindo uma seq. Shine Dalgarno (SD) e um códon de iniciação
- SD é localizada a 7 ± 2 nt *upstream* do códon de iniciação



É o AUG nos sistemas eficientes de expressão recombinante.

SD= UAAGGAGGG

A estrutura secundária do RBS é muito importante para a iniciação da tradução – a eficiência melhora com altos níveis de A e T.

A iniciação da tradução é influenciada pelo códon seguinte ao iniciador.

“A” é abundante em genes altamente expressos.

Estabilidade do mRNA

Os níveis de expressão gênica são principalmente determinados pela:

1. Eficiência da transcrição;
2. Estabilidade do mRNA;
3. Frequência da tradução.

- 1 e 3 têm sido otimizados, mas 2 muito pouco...
- A expressão do gene é controlada pelo decaimento do mRNA. A $\frac{1}{2}$ vida média do mRNA em *E. coli*, a 37° C, varia de segundos a 20 minutos.
- Causas da degradação do mRNA:
 - ataque por nucleases (RNase E)
 - falta / incompleto: enovelamento, proteção pelo ribossomo

O RNAm e a terminação da transcrição e tradução

- Terminadores de transcrição estabilizam o mRNA pela formação de um *stem loop* no 3'.
 - Em *E. coli*, a terminação da tradução é preferencialmente mediada pelo códon de parada “UAA”
- códons de parada consecutivos ou prolongados (UAAU) são mais eficientes.

Combinação de códons de terminação e fatores de liberação:

- RF-1 - UAG
- RF-2 - UGA
- RF-1 e RF-2 - UAA

Considerações sobre a linhagem hospedeira

Linhagens de expressão

- A linhagem para expressão é muito importante;

Linhagens de expressão deveriam ser deficientes na maioria das proteases naturais problemáticas e manter o plasmídeo de expressão estável.

- BL21 é o hospedeiro mais comum p/ expressão recombinante;

É robusta, mas não patogênica;

deficiente em proteases *ompT* e *lon*

Utilização de códonos (*Codon usage*)

- Presença de códonos raros na proteína de interesse...

Códonos raros em *E.coli*

AGA, AGG, CGA, CGG.....	Arg
UGU, UGC.....	Cys
GGA, GGG.....	Gly
AUA.....	Ile
CUA, CUC.....	Leu
CCC, CCU, CCA.....	Pro
UCA, AGU, UCG, UCC.....	Ser
ACA.....	Thr

Interferência de códons raros

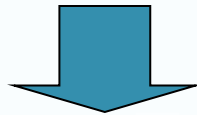
Códons preferenciais são reflexo do nível de tRNAs aminoacilados disponíveis no citoplasma.

- Códons preferenciais $\Rightarrow$ genes altamente expressos
- Códons raros $\Rightarrow$ expressos em baixos níveis.
- Códons raros para *E. coli* em genes heterólogos:
 - ausência de expressão; erros de tradução (substituição de aa; mudança de fase de leitura; terminação prematura);

PORTANTO:
a expressão da proteína em altos níveis não implica em integridade traducional!!

Estresse induzido pela super-expressão

- Algumas proteínas recombinantes influenciam diretamente o metabolismo celular;
- Mais comumente geram um aumento da carga metabólica.



Desviam recursos do metabolismo hospedeiro para manter a expressão recombinante

Genes *housekeeping* podem ser *down* regulados;

Falta de aminoácidos e/ou cofatores

Aumento da proteólise *in vivo*

Degradação proteolítica

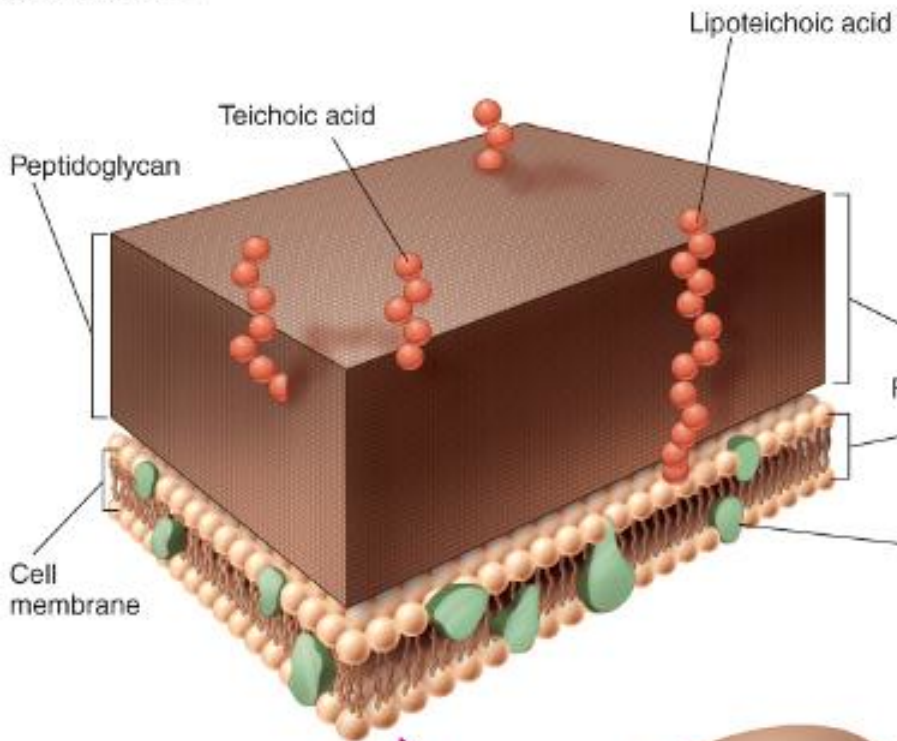
Estratégias para contornar o problema:

- Linhagens deficientes em proteases
- Proteínas de fusão
- Condições de crescimento (baixas temperaturas)
- Co-expressão com chaperonas
- Direcionamento da proteína para o meio extracelular ou espaço periplasmático
- Mutagênese eliminando os sítios de reconhecimento.

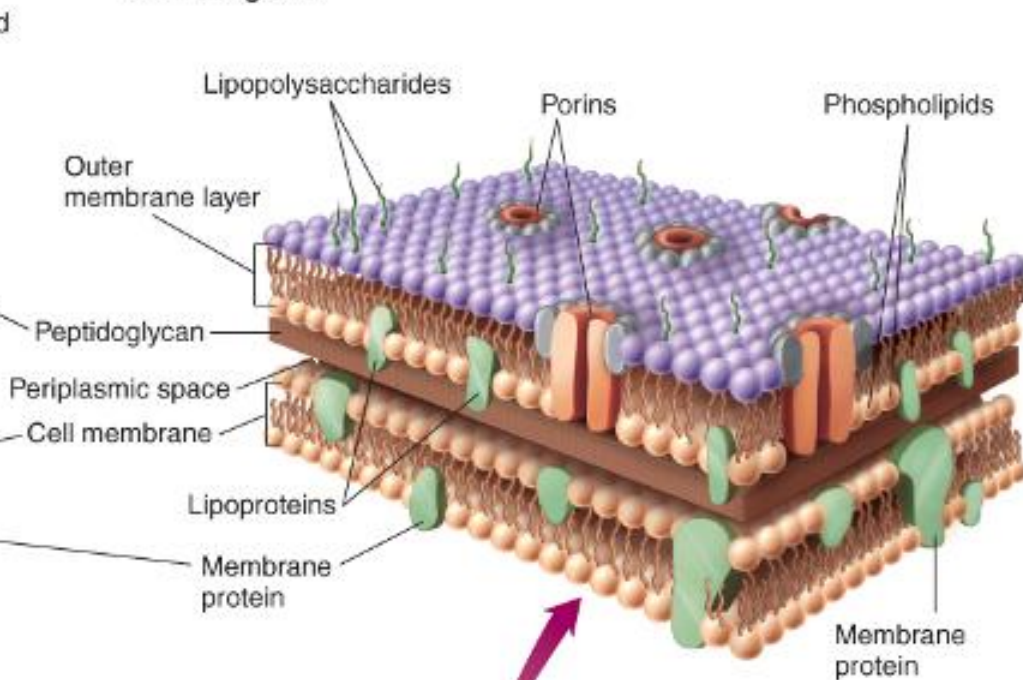
O destino das proteínas recombinantes:

Periplasma
Citosol
Secreção.

Gram-Positive



Gram-Negative



O destino das proteínas recombinantes

Citoplasma: alto rendimento

Tioredoxina e glutaredoxina reduzem Cys no citoplasma

Trx e glutathione redutases mantêm essas proteínas reduzidas. Bloqueio de *trx B* e *gor*, que codificam as redutases, facilita a formação de pontes S-S no citoplasma.



Linhagens disponíveis: AD494 (*trxB*-) Origami (*trx*- *gor*-)

Periplasma: catálise de pontes S-S, porém menor rendimento

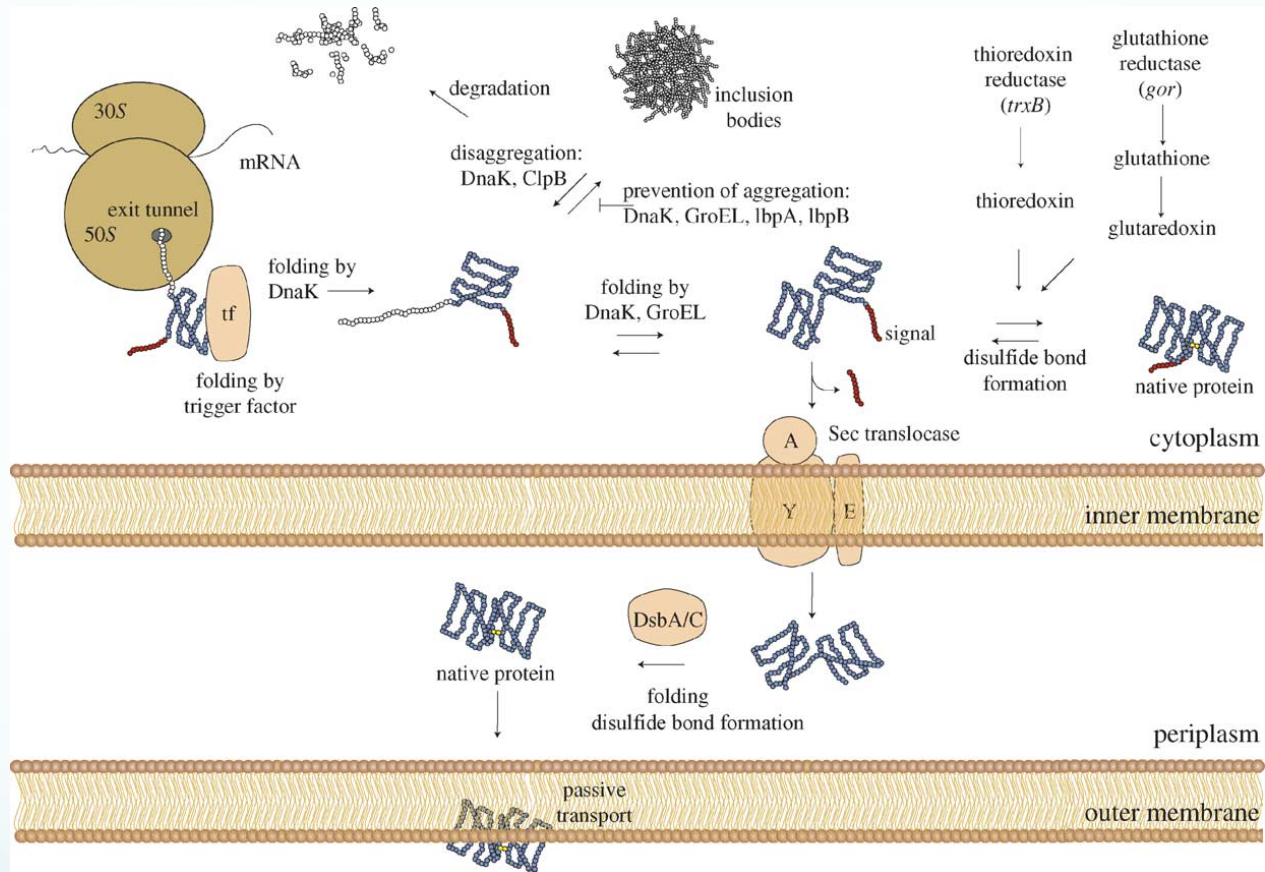
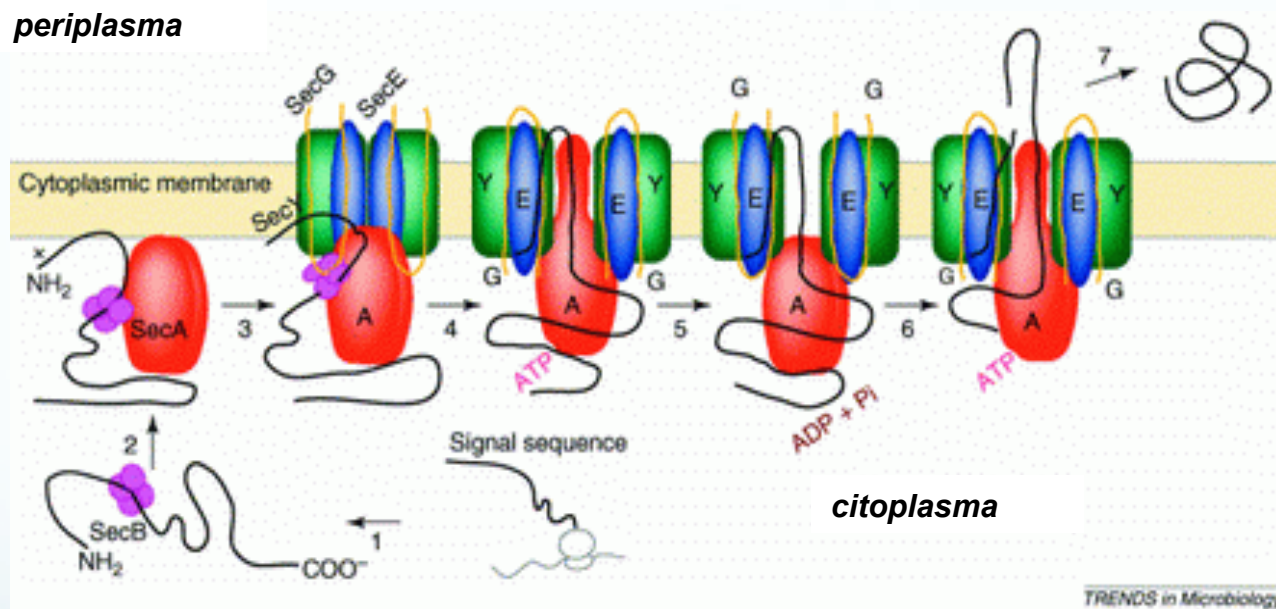


Fig. 3. Protein folding and secretion in *E. coli*. Pathways important for recombinant expression, secretion and disulfide bond formation are shown. See text for details and references.

Exportação e Secreção

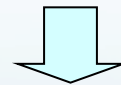
- Maior parte da proteína que é exportada vai p/ periplasma (via SEC)



Requer uma seqüência sinal (ex: *PhoA*, *OmpA*, *OmpT*, *LamB*, *OmpF*)

Secreção em *E. coli*

- Para o meio extracelular:
vias eficientes de translocação através da membrana externa são ausentes.
 - Requer sequência sinal (ex: hemolisina, pelB)
- algumas proteínas podem se difundir a partir do periplasma p/ o meio.
- Transporte passivo via membrana “desestabilizada”
cél. velhas ou que co-expressam fatores de secreção



Sec *E* e Sec *Y*

Sistemas que co-expressam múltiplos alvos

- Por ex. proteínas hetero-oligoméricas:

- co-expressão em plasmídeos separados;

Para a manutenção em *E. coli*, replicon deve ser compatível.

- co-expressão em plasmídeos policistrônicos
permite a expressão de mais de dois genes.

Duet™ Vectors let you express up to 8 proteins in the same cell

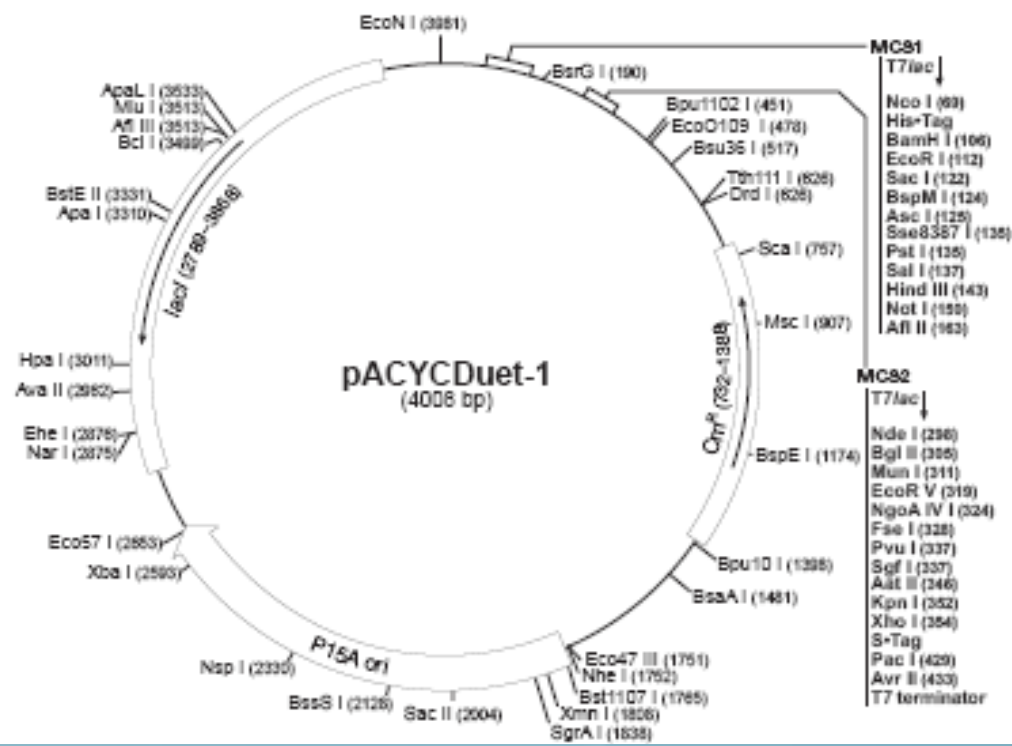


pACYCDuet-1 Vector

TB326 10/02

	Cat. No.
pACYCDuet-1 DNA	71147-3
pACYCDuet-1 sequence landmarks	
T7 promoter-1	2992-4008
T7 transcription start-1	1
His-Tag [®] coding sequence	83-100
Multiple cloning sites-1 (<i>Nco</i> I- <i>Afl</i> II)	69-168
T7 promoter-2	214-230
T7 transcription start-2	231
Multiple cloning sites-2 (<i>Nde</i> I- <i>Avr</i> II)	297-438
S-Tag [™] coding sequence	366-410
T7 terminator	462-509
P15A origin	1750-2062
<i>cat</i> (<i>Cm^R</i>) coding sequence	732-1388
<i>lacI</i> coding sequence	2789-3868

pACYCDuet™-1 is designed for the coexpression of two target genes. The vector contains two multiple cloning sites (MCS), each of which is preceded by a T7 promoter/*lac* operator and ribosome binding site (rbs). The vector also carries the P15A replicon, *lacI* gene and chloramphenicol resistance gene. This vector can be used in combination with pETDuet™-1 (Cat. No. 71146-3) in an appropriate host strain for the coexpression of up to 4 target genes. Genes inserted into MCS1 can be sequenced using the ACYCDuetUP1 Primer (Cat. No. 71178-3) and DuetDOWN1 Primer (Cat. No. 71179-3). Genes inserted into MCS2 can be sequenced using the DuetUP2 Primer (Cat. No. 71180-3) and T7 Terminator Primer (Cat. No. 69337-3).



Sistemas de Expressão Procariontes Baseados na RNAP T7

Sistemas de expressão baseados na RNAP T7

- Plasmídeos com o PT7 eram transcritos quando a cél. hospedeira era infectada com o fago T7.



A RNAP gerada pelo fago agia *in trans* no PT7.

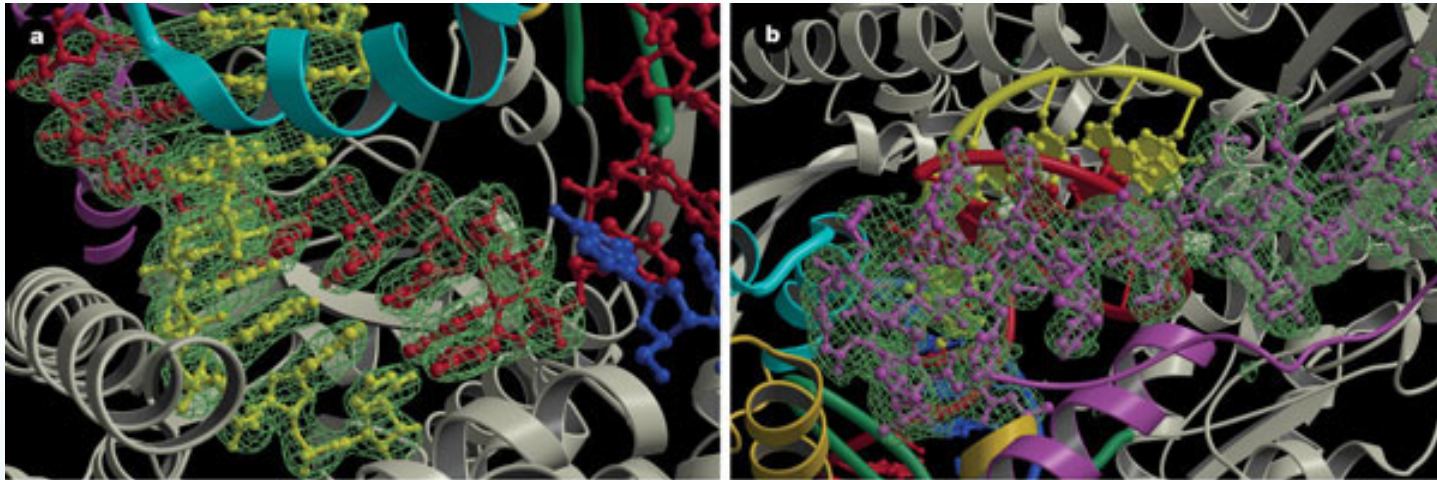
- Studier *et al.* (1990) desenvolveram sistemas de expressão usando o *gene 1* do fago T7: série **pET**



plasmídeos p/ **E**xpressão pela RNAP do **T7**

- O *gene 1* codifica a RNAP do T7, que é altamente específica p/ o PT7 e assim “dirigirá” a transcrição dos genes sob o controle desse promotor.

Estrutura da RNAP do T7 a 2,9Å (*Nature*, 2002)



Sistemas de expressão baseados na RNAP T7

O sistema de expressão porém fica restrito ao “fornecimento” de RNAP T7 exógena.

Como fazer isso?

- Infecção da linhagem hospedeira pelo fago T7...

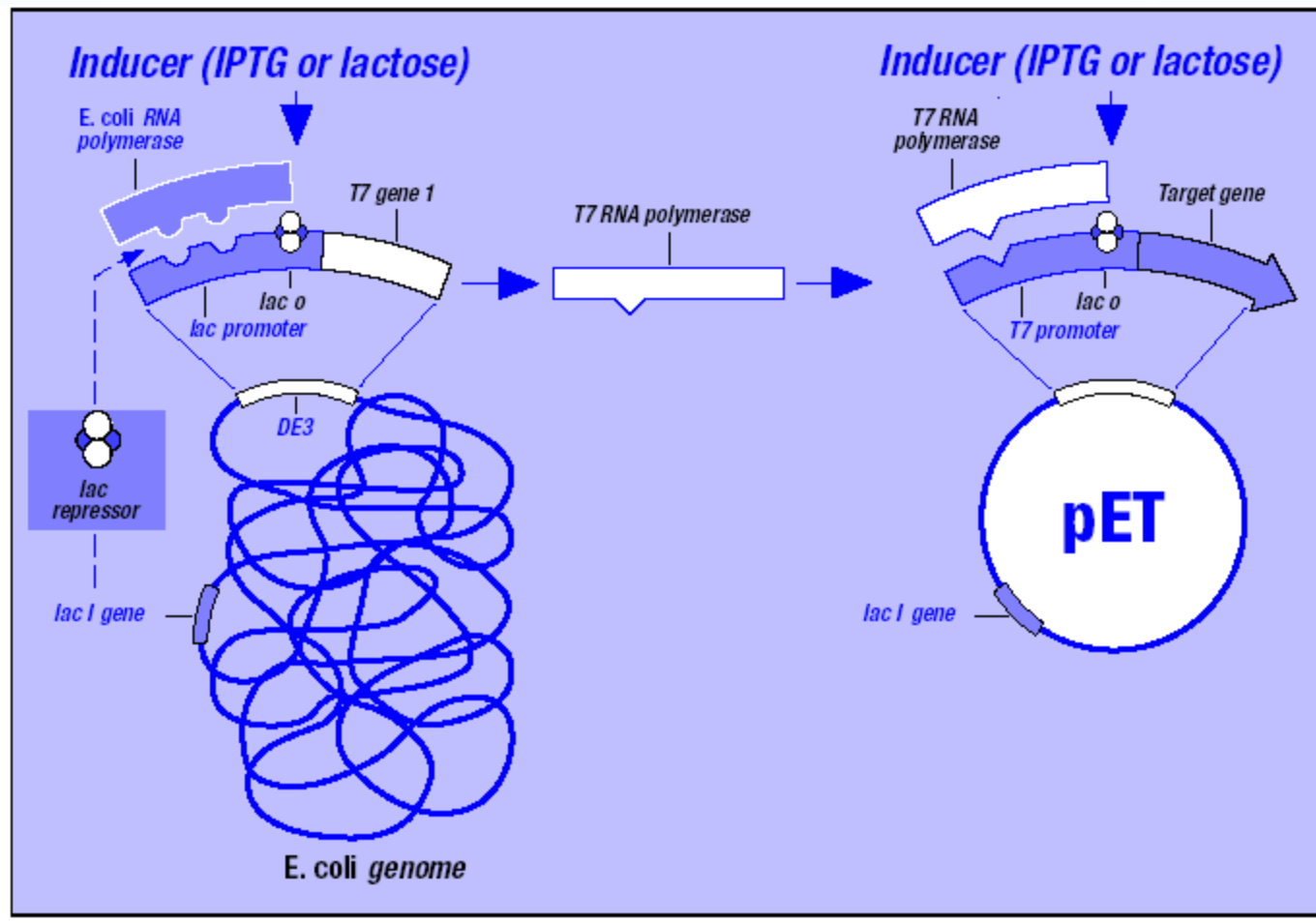
Morte celular em poucos minutos...

- Usar uma linhagem de *E. coli* que sintetize a RNAP T7

Por ex., BL21(DE3)



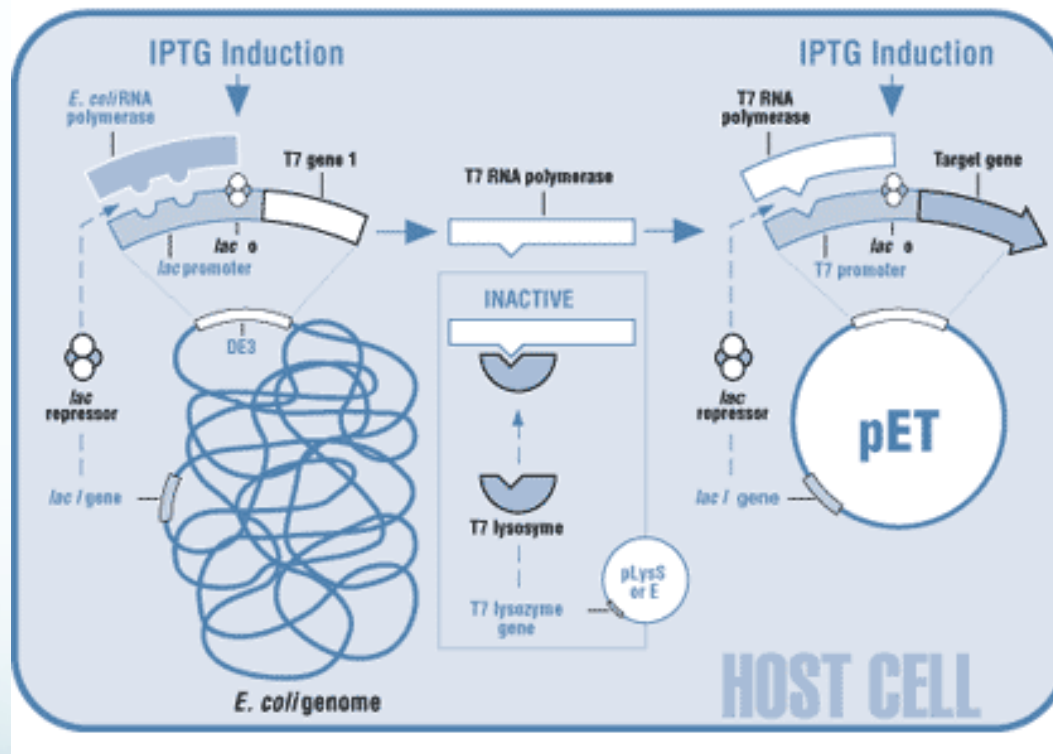
gene 1 do T7 inserido no genoma de BL21



A linhagem BL21

- deficiente em:
 - protease de membrana *ompT* (“cliva” a RNAP T7)
 - protease citoplasmática *lon1*.
- O *gene 1* em BL21 (DE3) está sob o controle do promotor *lacUV5* → Induzível por IPTG
- Desvantagem do sistema: nível basal de expressão da RNAP T7 ocorre na ausência de indução...

Controle da expressão basal



BL21(DE3) pLysS e pLysE

Lisozima do T7 : um inibidor natural da RNAP T7

Proteína bifuncional →liga-se a RNAP T7 ,
inibindo a transcrição



“cliva” uma ligação específica da parede cel. de *E. coli*

Altos níveis endógenos são tolerados em *E. coli*

pLysS e pLysE são compatíveis com os vetores pET

Rendimento

Varia com:

- a proteína expressa x sistema usado

Rendimentos usando o PT7 em *E. coli* variam de

- Dezenas a centenas de mg/litro de cultura.
- A proteína alvo pode consistir de até 60% da proteína total bacteriana após a indução.

Isso pode ser às custas da insolubilidade...

Leitura complementar:

“Expression in *Escherichia coli*: becoming faster and more complex” in *Current Opinion in Structural Biology* 2013, 23:326–334