

**PRODUZIONE DI PROTEINE
RICOMBINANTI**

=

**UNO DEI PRINCIPALI INTERESSI
DELLE INDUSTRIE**

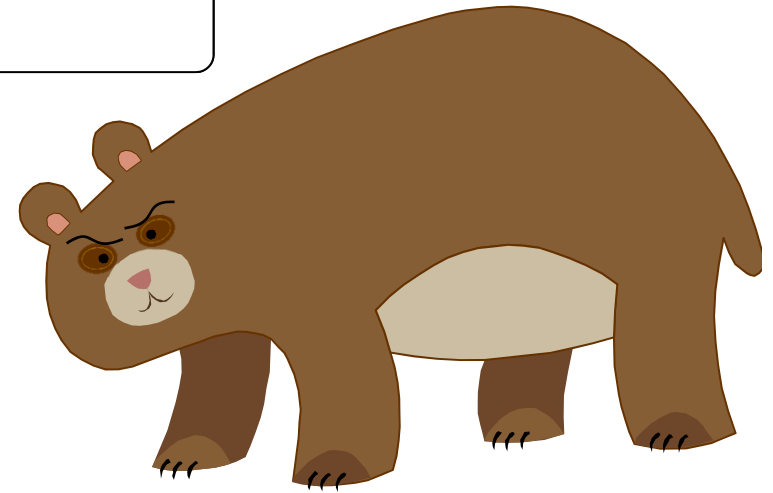
**NON
RICOMBINANTI**

**DIFFICOLTA' DI
OTTENERE LE FONTI**

**SCARSITA' DEL
PRODOTTO**

**DIFFICOLTA' DI
PURIFICAZIONE**

**POSSIBILE PERICOLOSITA'
INTRINSECA**



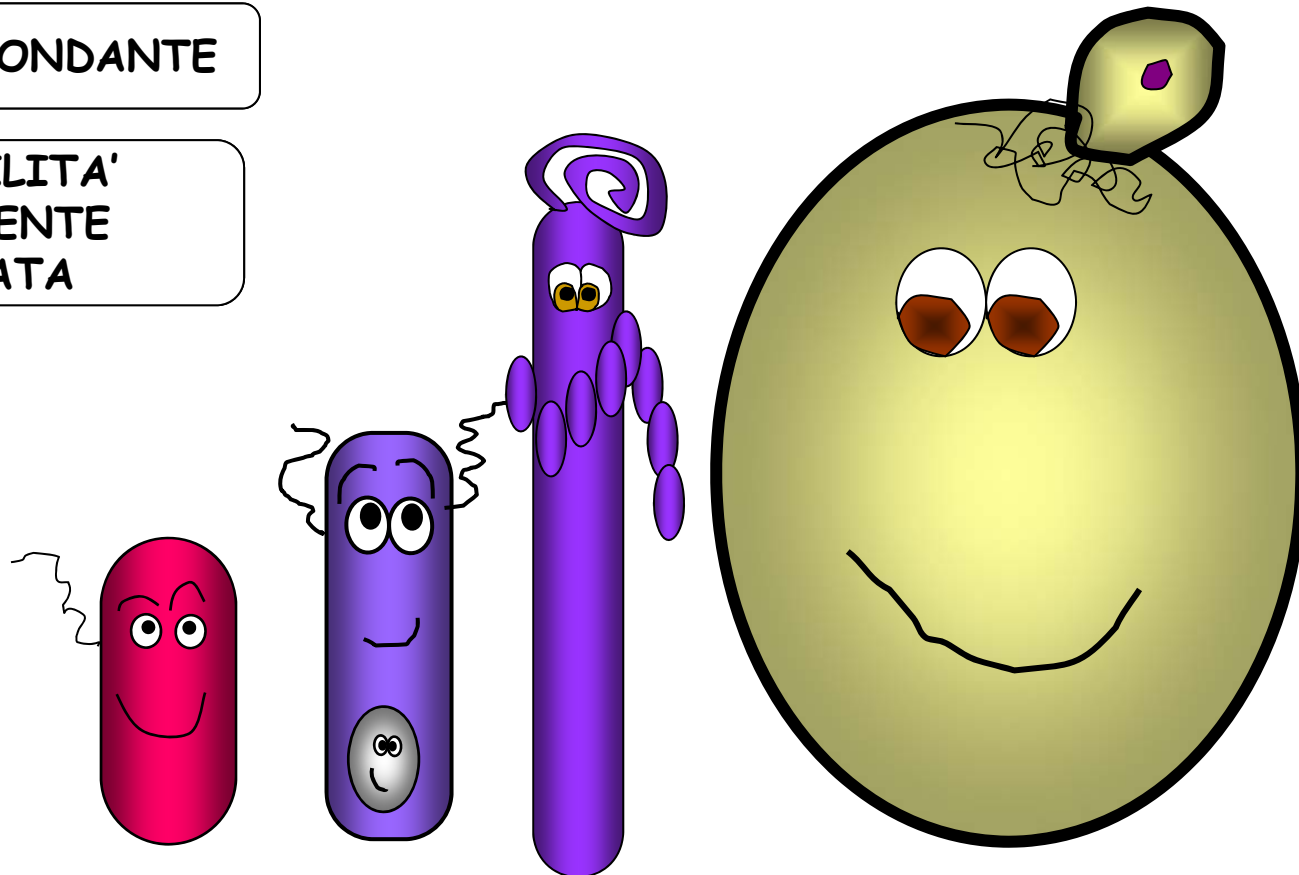
RICOMBINANTI

FACILE
PURIFICAZIONE

RESA ABBONDANTE

DISPONIBILITA'
PRATICAMENTE
ILLIMITATA

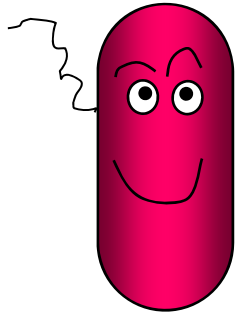
SICUREZZA



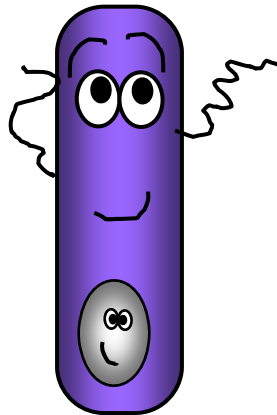
OSPITI-1

BUONA CONOSCENZA DELLA
GENETICA E DELLA FISIOLOGIA

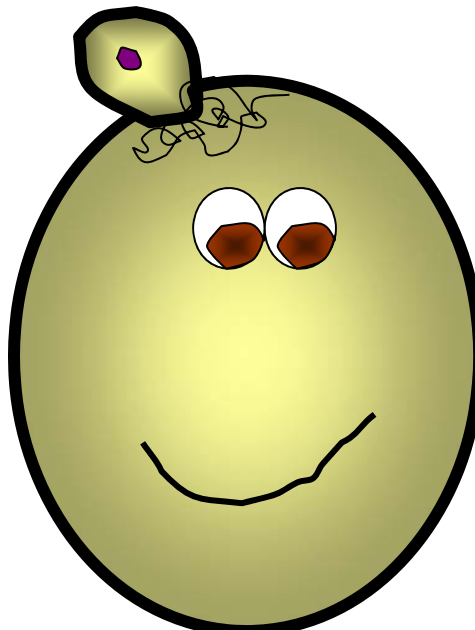
AMPIA DISPONIBILITA' DI VETTORI
E SISTEMI DI ESPRESSIONE



Escherichia coli



Bacillus subtilis

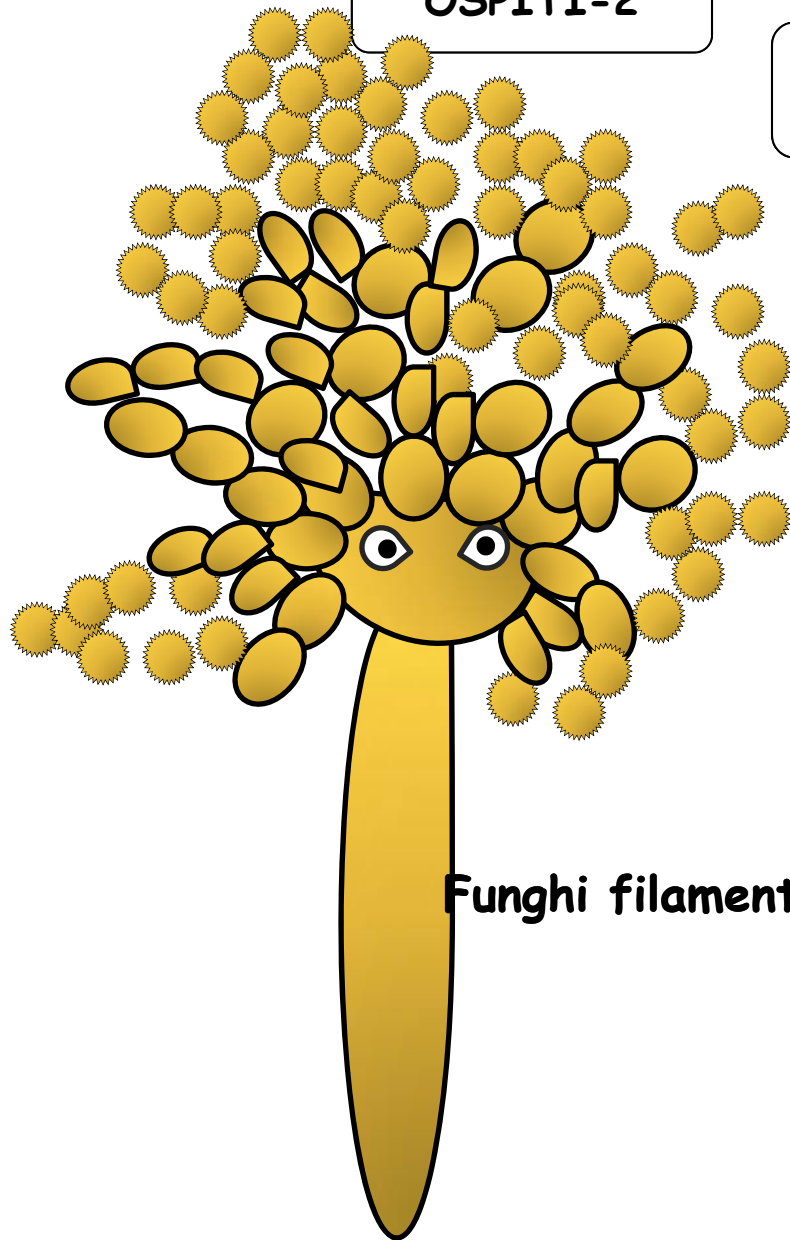


Saccharomyces cerevisiae

OSPITI-2

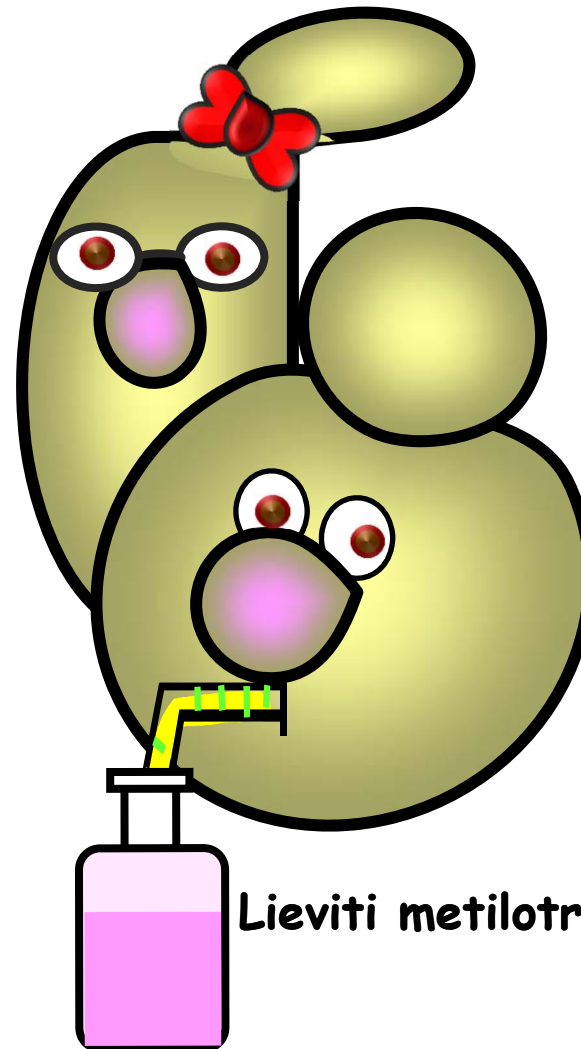
**CARATTERISTICHE
INTERESSANTI**

streptomiceti

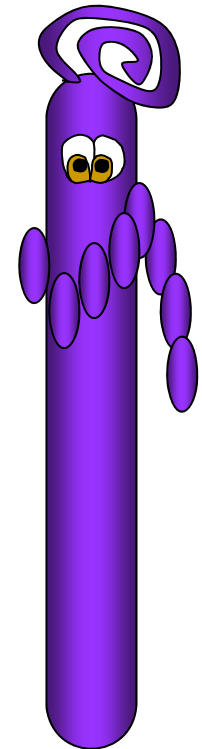


Funghi filamentosi

**MINORE DISPONIBILITA'
DI SISTEMI**



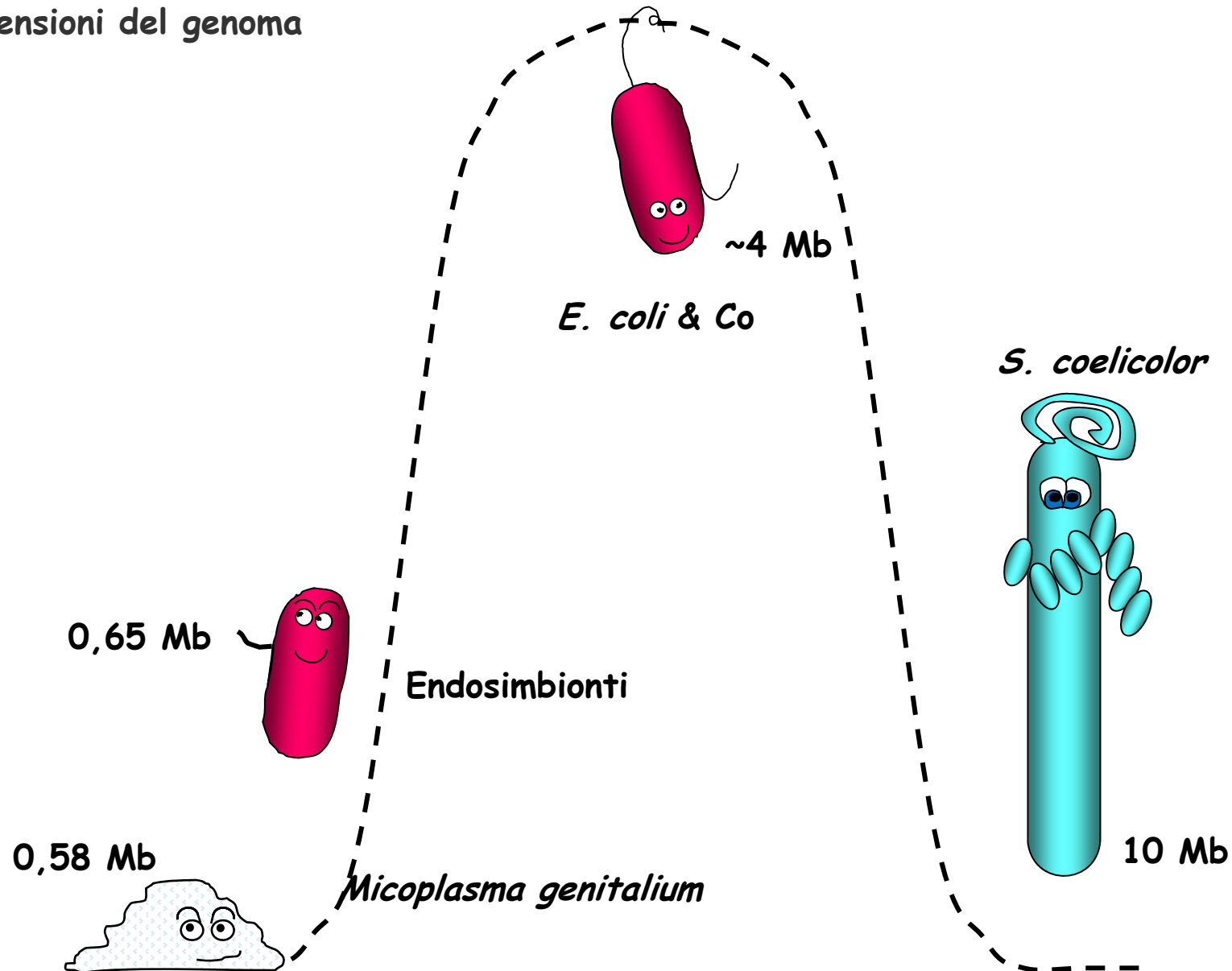
Lieviti metilotrofi



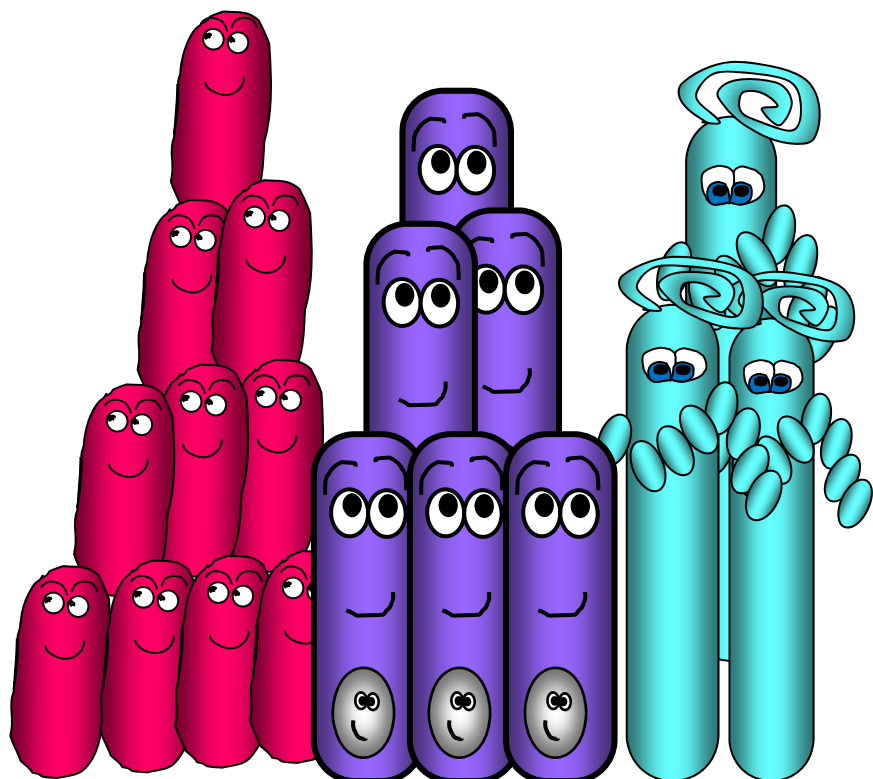
BATTERI: VANTAGGI

Aploidia

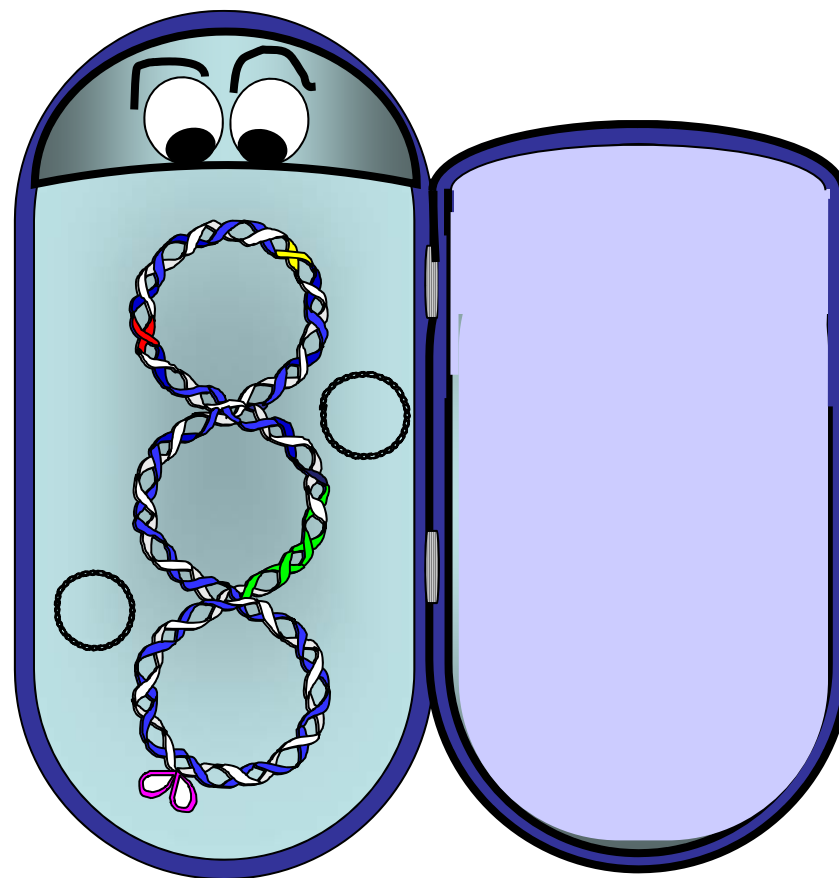
Dimensioni del genoma



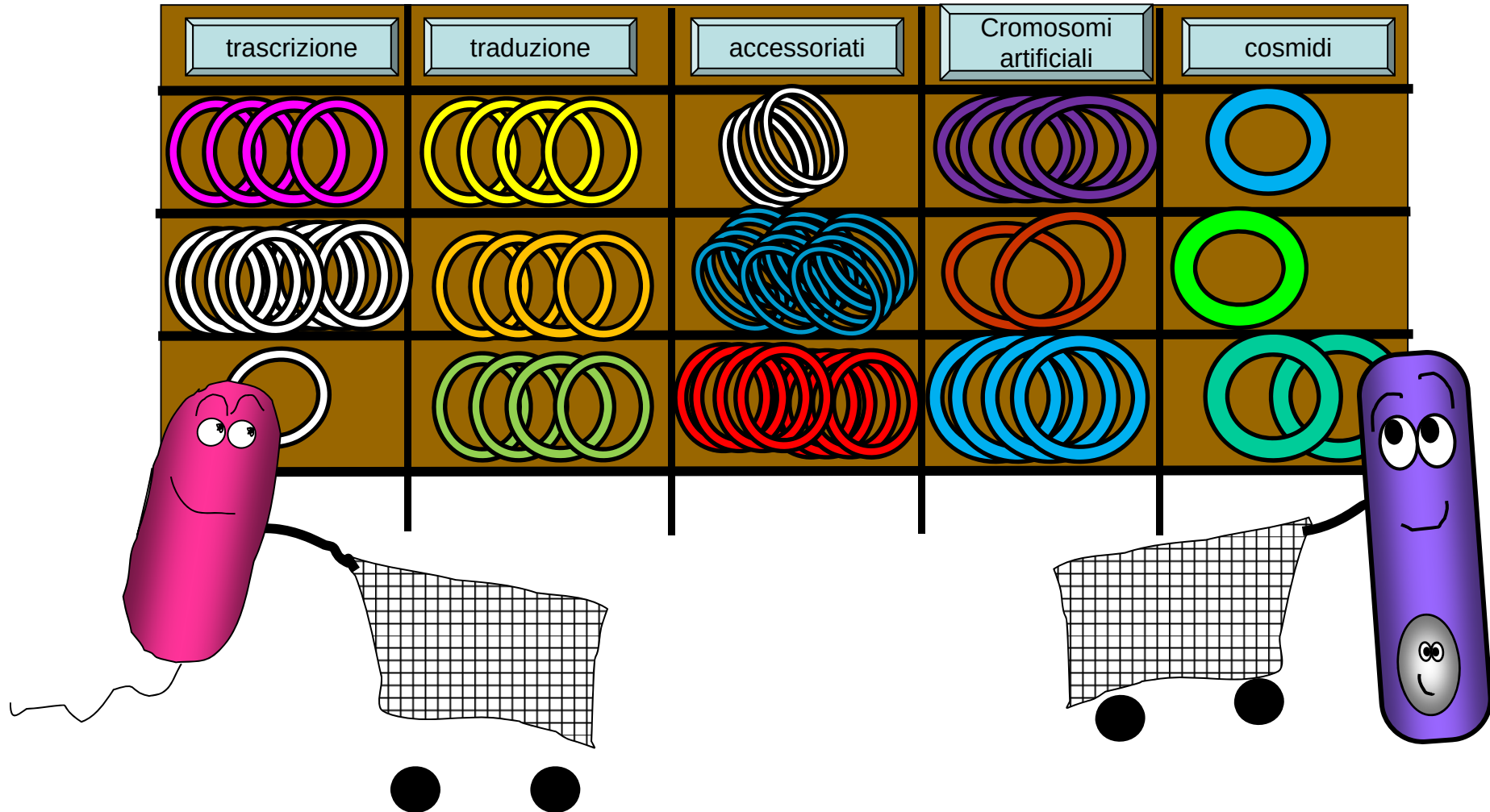
Ritmo di crescita



Facilità di manipolazione



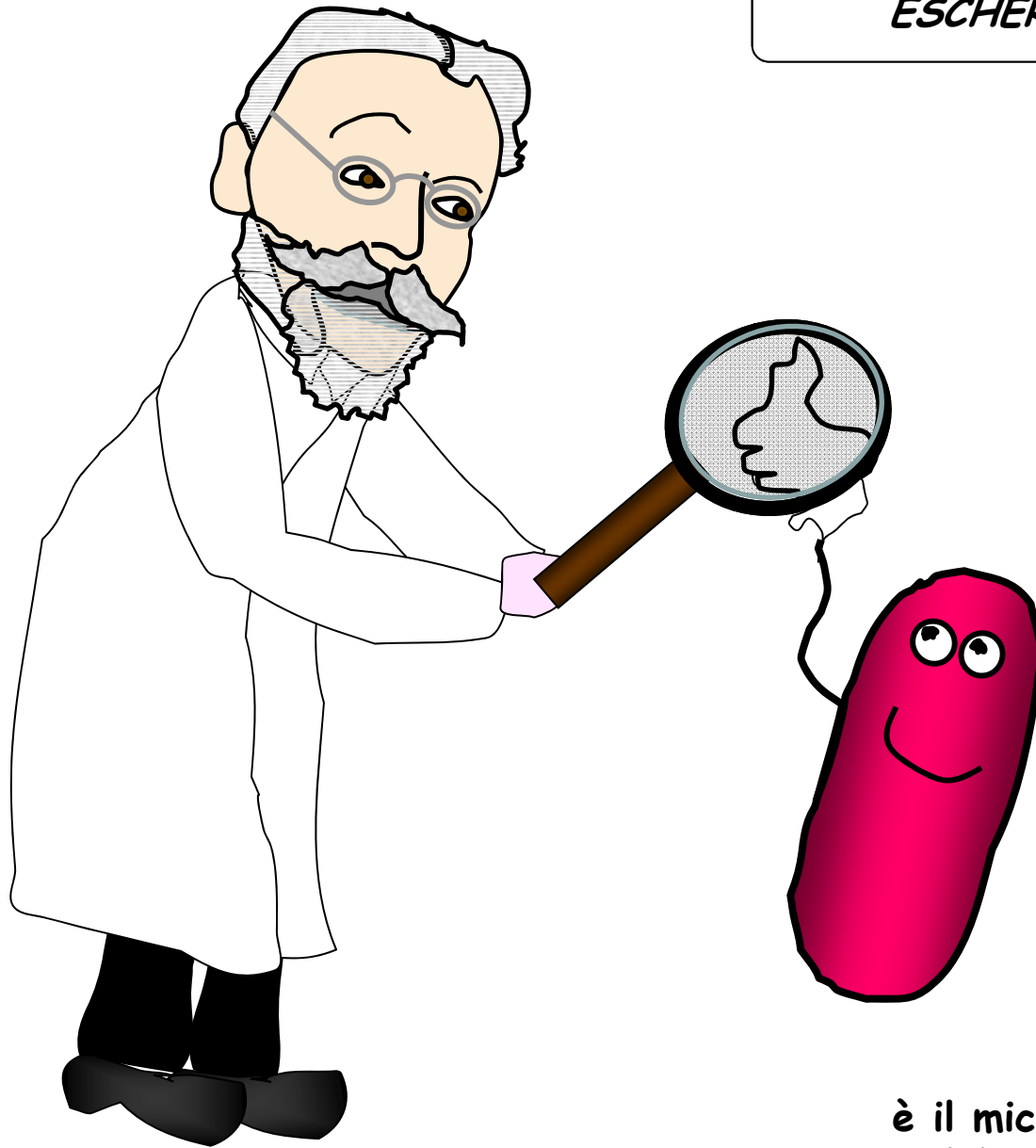
Abbondanza di plasmidi vettori e sistemi di espressione



SVANTAGGI: assenza di modificazioni delle proteine dopo la traduzione

ESCHERICHIA COLI

Identificato nel 1885
da Theodor Escherich



è il microrganismo più utilizzato nei
laboratori di ricerca e per la
produzione di proteine eterologhe

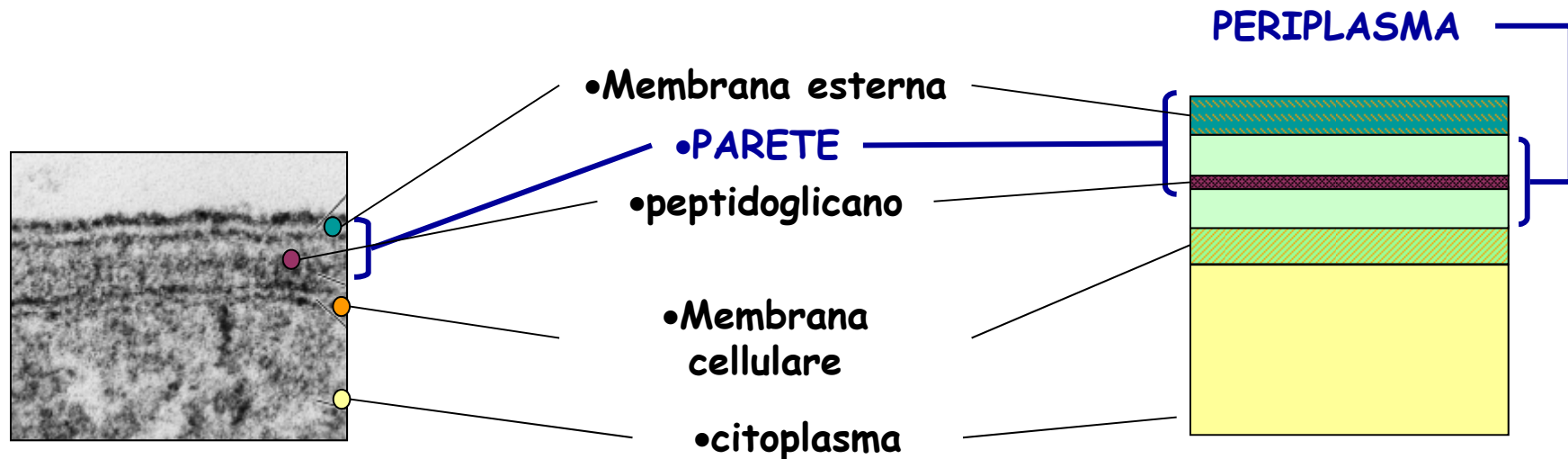
Batterio Gram negativo

proteobatteri

γ -proteobatteri

Enterobacteriaceae

Struttura didermica



PERIPLASMA

comparto ossidante, ~ 4% delle proteine totali

CITOPLASMA

comparto riducente

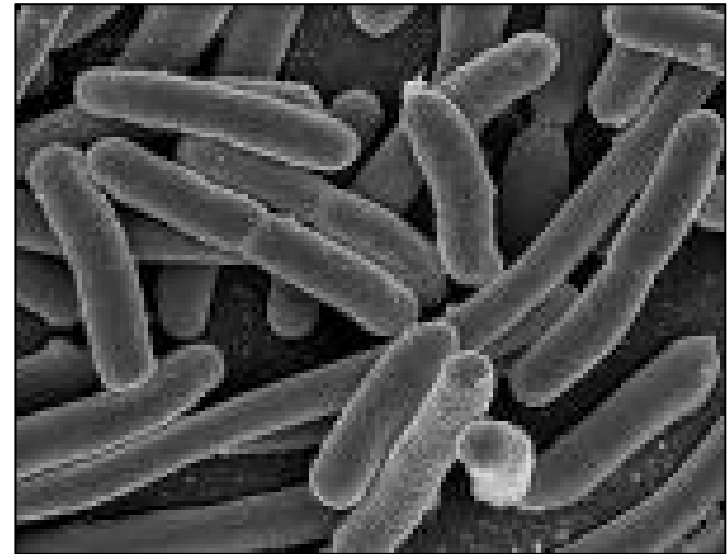
MEMBRANA ESTERNA

Delimita la cellula; può essere sfruttata per esprimere epitopi superficiali

bastoncello dritto, $\sim 2 \mu\text{m} \times 0,5 \mu\text{m}$

4 principali gruppi filogenetici
(A, B1, B2, D)

Moltissimi sierotipi



O (>170) = catene
polisaccaridiche LPS

Formula antigenica O: ... H... K...

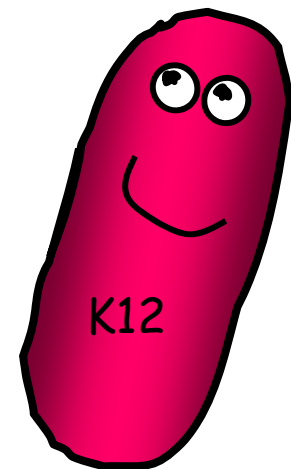
H (>50) = proteine
flagellari

K (>100) = capsulari
(equivalente microcapsulare)

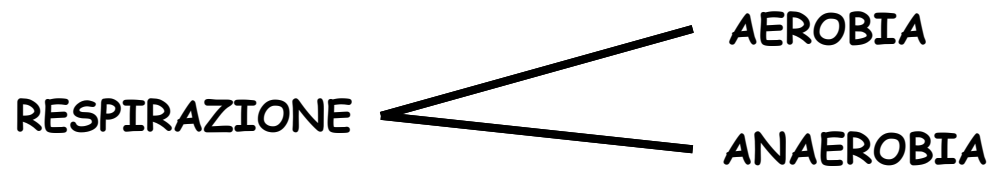
Escherichia coli K-12

Organismo modello

Strumento per
biotecnologie



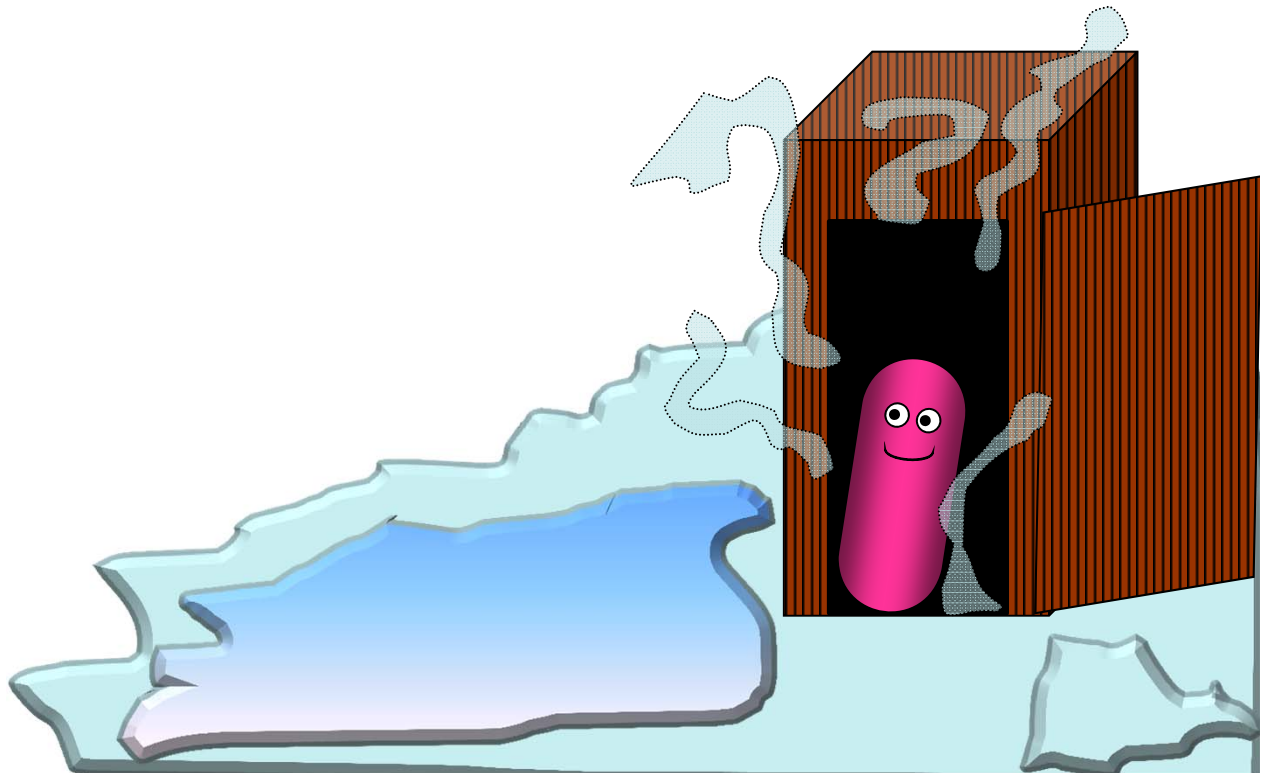
PRODUZIONE DI ENERGIA

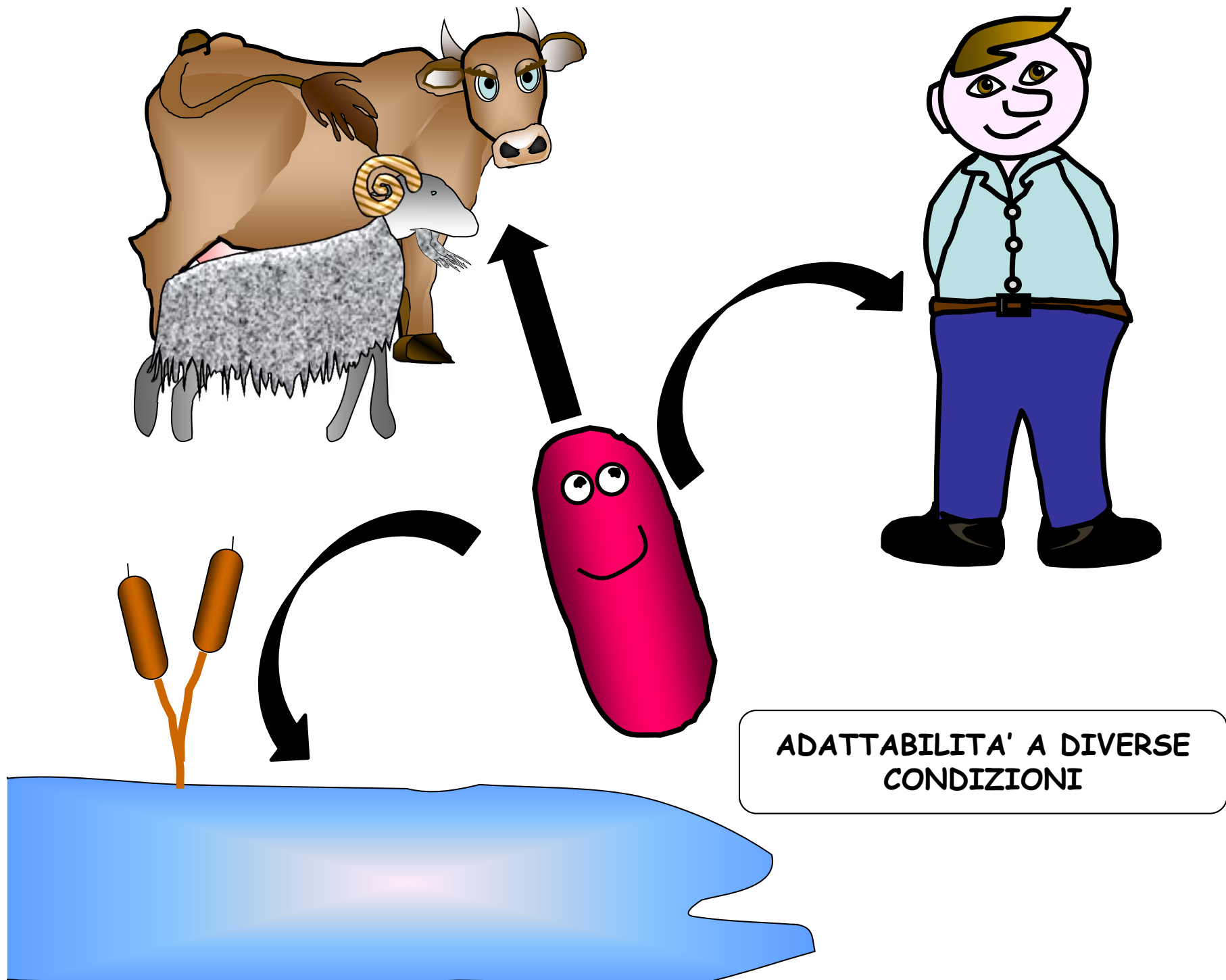


FERMENTAZIONE

TERRENI RICCHI (37 -42 °C)
TEMPO GENERAZIONALE MEDIO = 20'

SE NECESSARIO CRESCE
TRA 10-45 °C





E. coli come strumento

Genetica nota



Grande varietà
di mutanti

Biochimica delle
fermentazioni nota



Facilita l'allestimento di
processi industriali controllati

Buona competenza
Facile da ottenere

Crescita abbondante in
molti terreni

Biomassa abbondante, le
cellule si lisano facilmente

Classe di patogenicità = 1



Sicurezza d'uso

POSSIBILI SVANTAGGI (*E. coli*)

**Mancanza di modificazione
(comune ai batteri)**

LPS



**Prodotti per uso umano: necessaria
ulteriore purificazione**

**Problemi di resa
(occasional)**

USO DEI CODONI

**CORPI DI
INCLUSIONE**

**STRESS
METABOLICO**

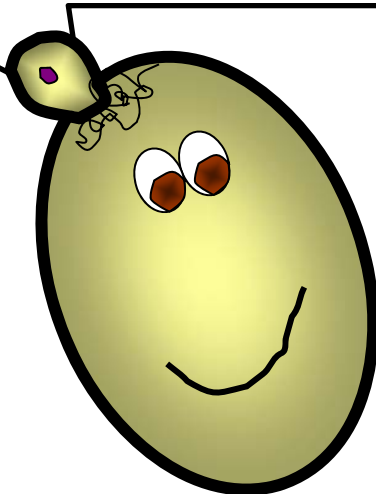
Il significato dei codoni
non è sempre uguale

E la frequenza d'uso varia
da organismo a organismo

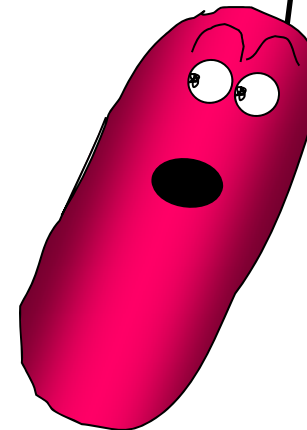
Alcuni codoni non universali

Codon	Codice universale	mitocondri (mammifero)	mitocondri (lievito)
UGA	Stop	Triptofano	Triptofano
AUA	Isoleucina	Metionina	Metionina
CUA	Leucina	Leucina	Treonina
AGA-AGG	Arginina	Stop	Arginina

TCGGCCCTAGTCGAG*GATAT
A*TGTCATGAG*TCGCCTGAG
TGTATAGAG*TCGGAGATA?



TCTGCTCTGGTTGAA*GACA
TC*TGCCATGAA*TCTCCGGA
ATGCATCGAA*TCTGAAATC?

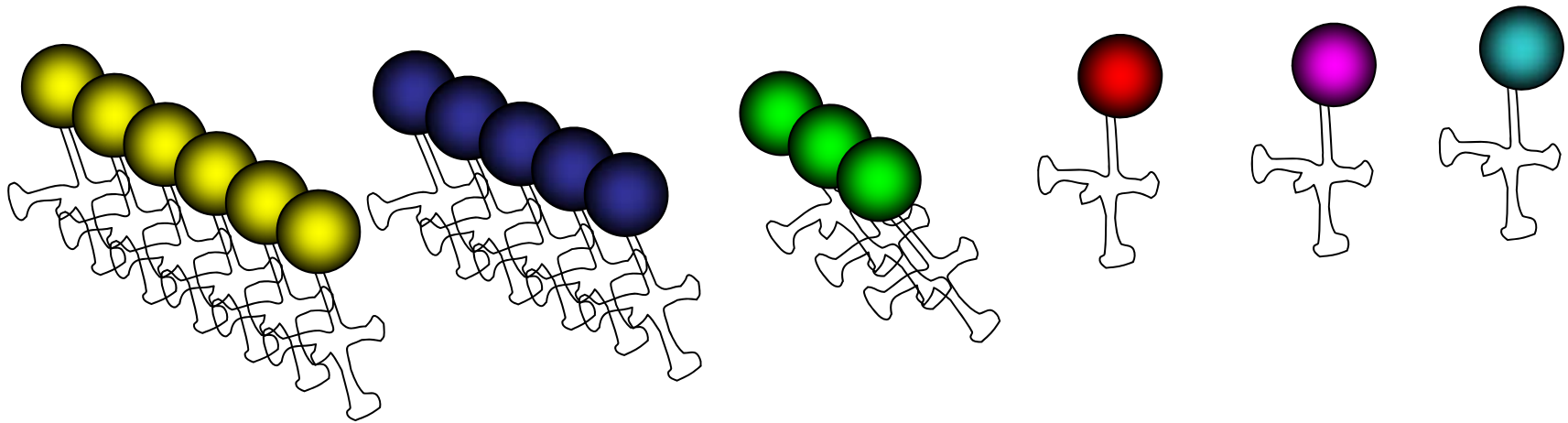


Organismo	ARG (AGG)	ARG (AGA)	LEU (CUA)	ILE (AUA)	PRO (CCC)	GLY (GGA)
<i>E. coli</i> B	2,1	2,4	3,4	5,0	2,4	8,2
<i>E. coli</i> K12	1,2	2,1	3,9	4,3	5,5	7,9
<i>Anabaena</i>	2,6	8,3	14,0	8,3	13,0	12,4
<i>B.megaterium</i>	2,7	9,1	10,9	10,3	2,9	26,2
<i>B.subtilis</i>	3,9	10,5	4,8	9,3	3,3	21,8
<i>Caulobacter</i>	2,2	0,8	1,4	0,6	18,7	4,3
<i>S.carnosus</i>	0,4	7,9	4,6	9,8	0,8	16,7
<i>S. lividans</i>	4,0	1,1	0,6	0,8	22,7	6,5
<i>Arabidopsis</i>	10,9	18,9	9,9	12,6	5,3	24,2
<i>Caenorhabditis</i>	4,0	15,4	7,9	9,4	4,4	31,6
<i>C. tetani</i>	5,0	25,5	11,2	67,4	1,8	34,6
<i>Drosophila</i>	6,3	5,2	8,2	9,5	18,0	17,7
<i>Homo sapiens</i>	11,9	12,0	7,2	7,4	19,9	16,5
<i>H. sapiens (mit)</i>	0,4	0,4	70,4	44,5	33,9	18,9
<i>Plasmodium</i>	3,3	17,0	5,3	40,8	3,0	20,0
<i>Pichia pastoris</i>	6,6	20,2	10,9	11,7	6,7	19,1
<i>S. cerevisiae</i>	9,3	21,3	13,4	17,8	6,8	10,9

Una frequenza d'uso > 15/1000 nel frammento da esprimere, può rendere difficile l'espressione in *E. coli*

Le probabilità di riuscire a iper-esprimere un gene, dipendono anche dalla coerenza dell'uso dei codon tra il frammento estraneo e l'ospite

La composizione del pool di tRna in una cellula riflette l'uso dei codoni e quelli rari rappresentano un limite alla traduzione del prodotto genico

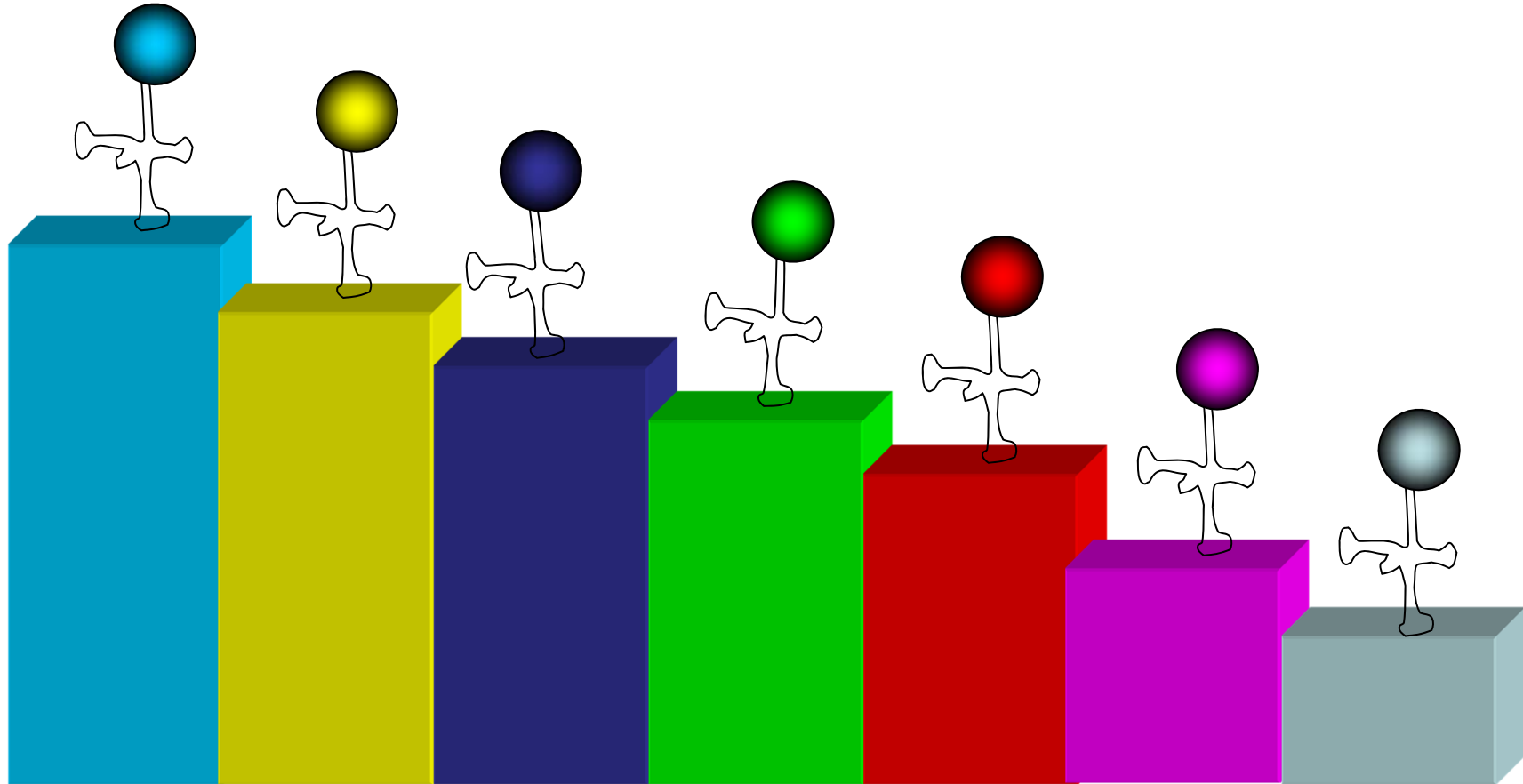


Provocando lo stallo nello scorrimento del ribosoma, terminazione precoce della traduzione, salti di frame e mutazioni per erranee incorporazioni di aa

al momento di scegliere un ospite bisogna valutare questo aspetto per scegliere l'ospite migliore

Per avere un'idea approssimativa delle probabilità di espressione si possono usare metodi matematici

Il metodo maggiormente usato è quello proposto nel 1986 da Sharp e Lee il "Codon Adaptation Index" (CAI)



Che assegna a tutti i singoli codon un preciso peso statistico valutato in base alla frequenza d'uso in un determinato ospite

Il peso statistico viene espresso con un valore (rscu),

codon	RSCU
AGA	0,01
AGG	0,008
CGA	0,017
CGC	1,561
CGG	0,017
CGT	4,38

Es: RSCU per l'aminoacido Arginina in *E. coli*

L'RSCU serve a determinare il singolo CAIobs
(osservato) per ogni aminoacido (CAIobsa)

elevando il suo valore ad un valore pari al numero di volte che quel
particolare codon viene usato nel gene che codifica la proteina in esame



Se in un peptide Ala è codificata 7 volte dal codon GCA

il cui valore RSCU è = 1,099

Il primo valore di CAIobs_(a) sarà $1,099^7 = 1,9363$

Altre 3 Ala sono codificate da GCC (0,228)

Il secondo valore cercato è $0,228^3 = 0,0118$

Il prodotto di tutte le potenze così calcolate

$CAIobs(a_1) \times CAIobs(a_2) \times \dots \times CAIobs(a_n)$

Elevato a $1/n^*$

è il "CAIobs" della proteina.

* Non si tiene conto degli aminoacidi
per i quali esiste un solo codon

Si calcola poi il CAI max (valore teorico) assumendo che ogni residuo nella proteina sia codificato dal proprio codon ottimale

Es.: 7 Alanine GCA+3 Ala GCC $\rightarrow 1,9363 \times 0,118 = 0,0228$

il migliore RSCU per Ala (1,877) è quello di GCT

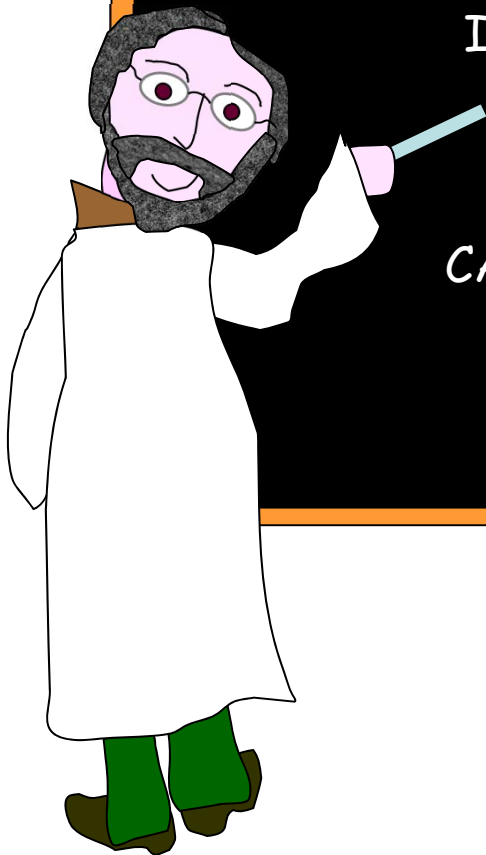
Il primo valore di CAI_{max}_(a) sarà $1,877^{(7+3)} = 542,80$

Il prodotto di tutte le potenze così calcolate

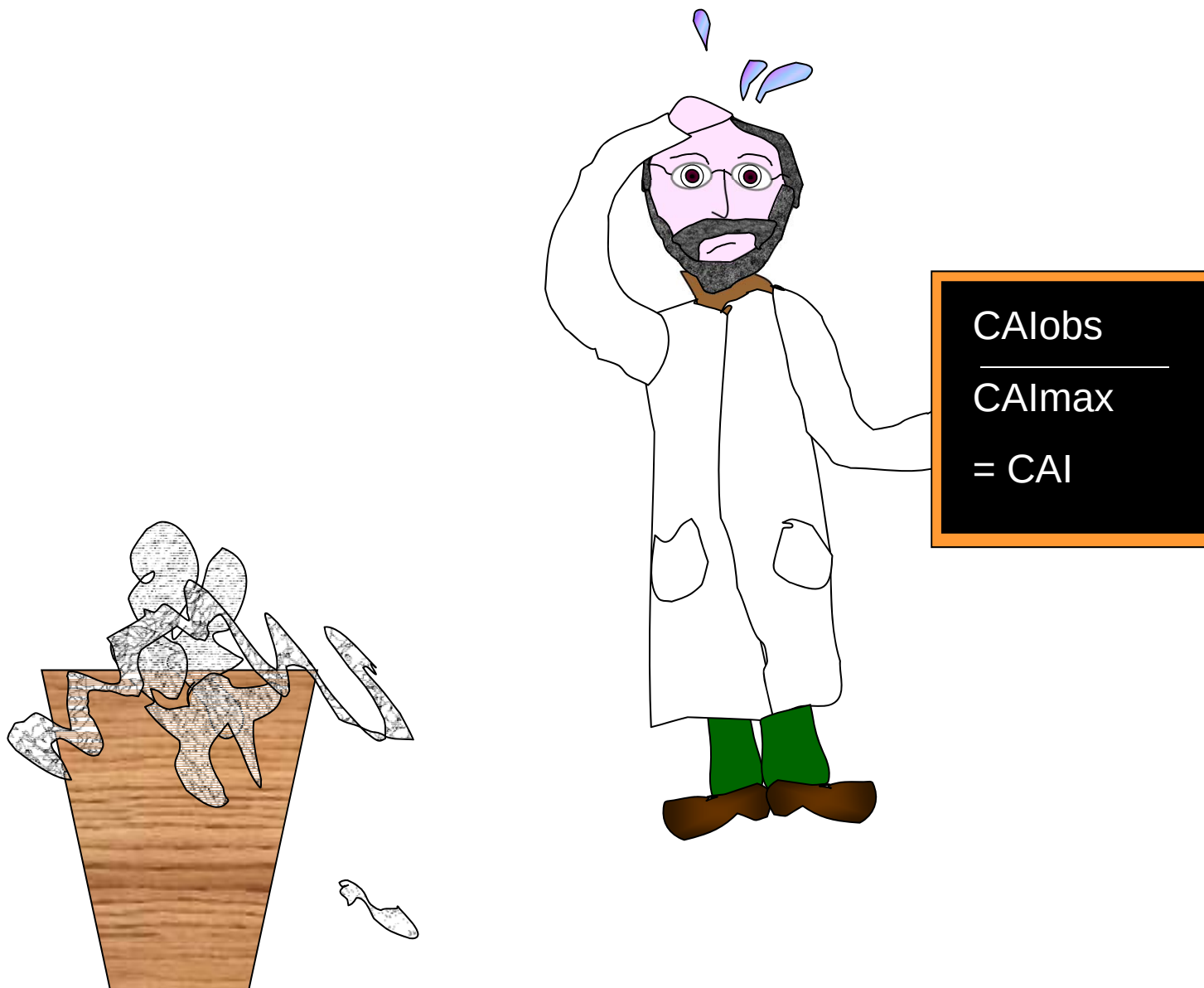
$CAI_{max}(a1) \times CAI_{max}(a2) \times \dots \times CAI_{max}(an) = CAI_{max}$

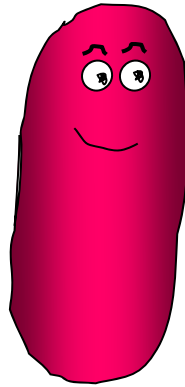
Elevato ad $1/n$

è il "CAI_{max}" della proteina.



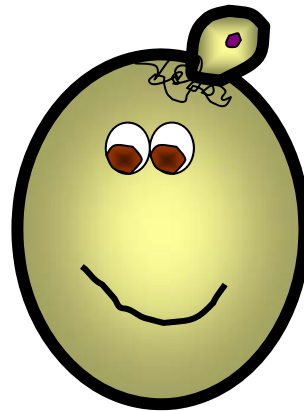
Infine, dividendo CAI_{obs}/CAI_{max} , si ottiene il valore di CAI .





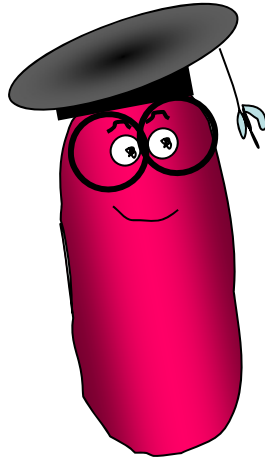
In *E. coli*

una proteina fortemente espressa ha CAI compreso tra 0,7 e 1
una proteina mediamente espressa ha CAI intorno a 0,5
e una proteina poco espressa ha CAI inferiore a 0,3



In *Saccharomyces*,

valori di CAI compresi tra 0,1 e 0,3 sono tipici di proteine regolative



RIEPILOGANDO:

CAI_{obs_a} = valore RSCU di un codon $\hat{n}^\circ$ di volte in cui è effettivamente usato

CAI_{obs} : moltiplicare tra loro tutte le potenze ottenute ed elevarle a $1/N$ (N =numero di residui analizzati)

CAI_{max_a} : VALORE max tra i codon per un aminoacido, $\hat{n}^\circ$ di volte TOTALE in cui quell'aminoacido compare nella proteina in esame

CAI_{max} : prodotto delle potenze ottenute con CAI_{max_a}

CAI : CAI_{obs} / CAI_{max}

Quanto più il valore di CAI si avvicina ad 1, tanto più si può prevedere che la proteina sarà espressa con facilità dal coli.

Prendiamo per esempio il Peptide "PERESPRIMERLASPLENDIDAMENTE"

Che può essere codificato dal frammento "A"

CCC	GAA	CGT	GAG	TCC	CCG	CGC	ATC	ATG	GAA	CGC	CTG	GCA	TCC
P	E	R	E	S	P	R	I	M	E	R	L	A	S
CCC	CTG	GAA	AAC	GAC	ATT	GAT	GCT	ATG	GAA	AAC	ACT	GAG	
P	L	E	N	D	I	D	A	M	E	N	T	E	

Oppure dal frammento "B"

CCC	GAG	AGG	GAG	TCG	CCC	AGA	ATA	ATG	GAA	AGG	TTA	GCC	TCG
P	E	R	E	S	P	R	I	M	E	R	L	A	S
CCT	TTG	GAA	AAT	GAT	ATA	GAT	GCC	ATG	GAG	AAT	ACA	GAG	
P	L	E	N	D	I	D	A	M	E	N	T	E	

I peptidi sono identici, ma calcoliamo il CAI dei due frammenti, usando gli RSCU di *E.coli*

$$\text{aminoacidi} = 27.. \quad \text{Met}=2; \text{Trp} = 0 \quad \rightarrow N = (27-2) = 25$$

$$1/N = 0,04 \quad \text{CAI}_{\text{max}} = 2,382602$$

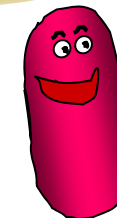
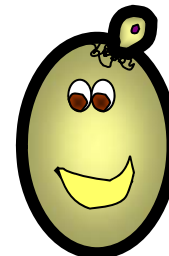
Usando la tabella degli RSCU si calcola che
CAI_{obs} del frammento (A) è = 1,188972
CAI_{obs} del frammento (B) è = 0,117304

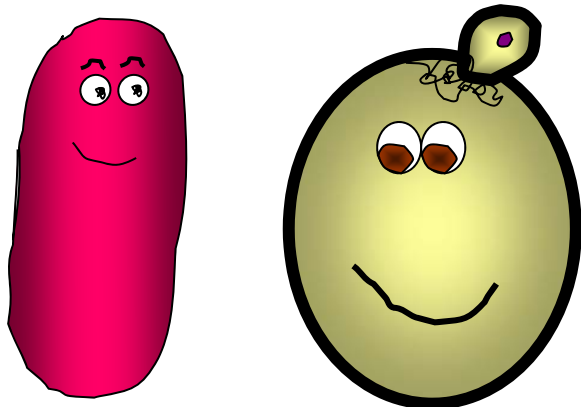
il CAI di (A) sarà $1,188972/2,382602 = 0,499$
il CAI di (B) sarà $0,117304/2,382602 = 0,0492$

10 volte di meno

una buona resa di PERESPRIMERLASPLENDIDAMENTE in *E. coli*, può essere ottenuta con "A"

Se si vuole usare il frammento (B),
bisognerà impiegare un ospite diverso

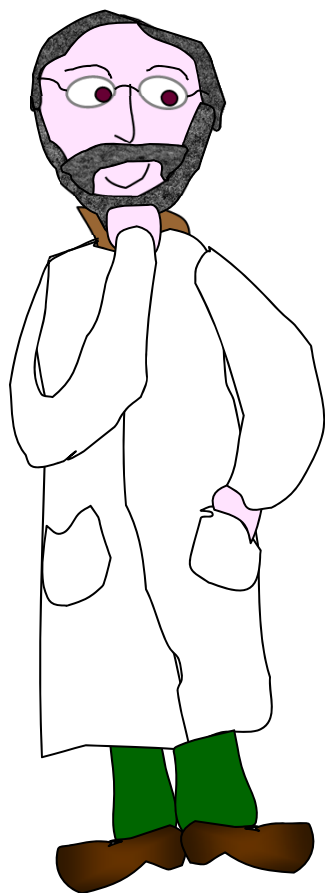




le tabelle complete degli RSCU sono disponibili
per *E. coli* e per *S.cerevisiae*.

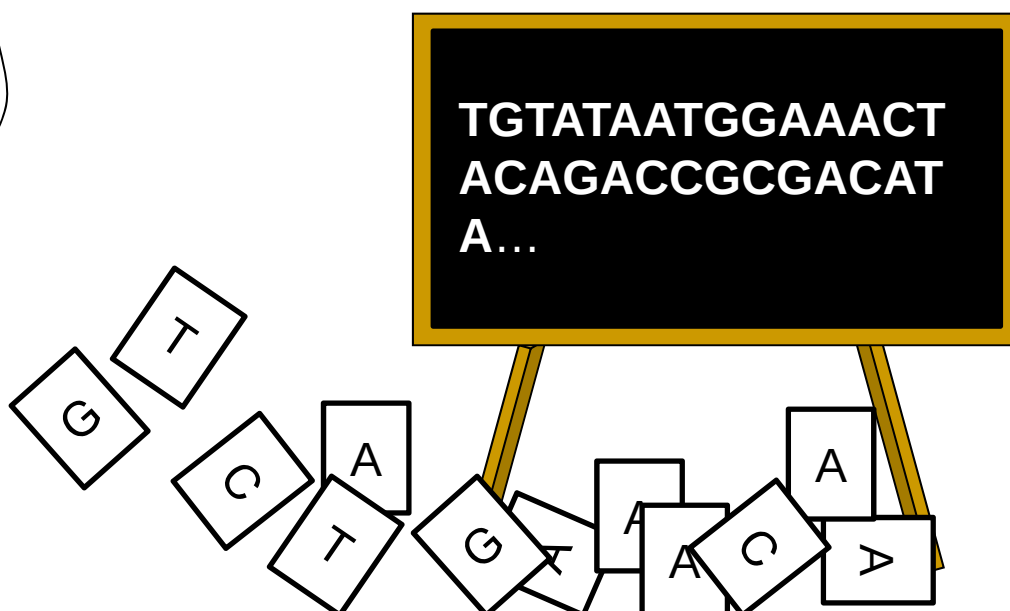
Ma il sequenziamento sistematico dei genomi ha reso
disponibili molti dati sulle frequenze di uso

(<http://www.kazusa.or.jp/codon/>)

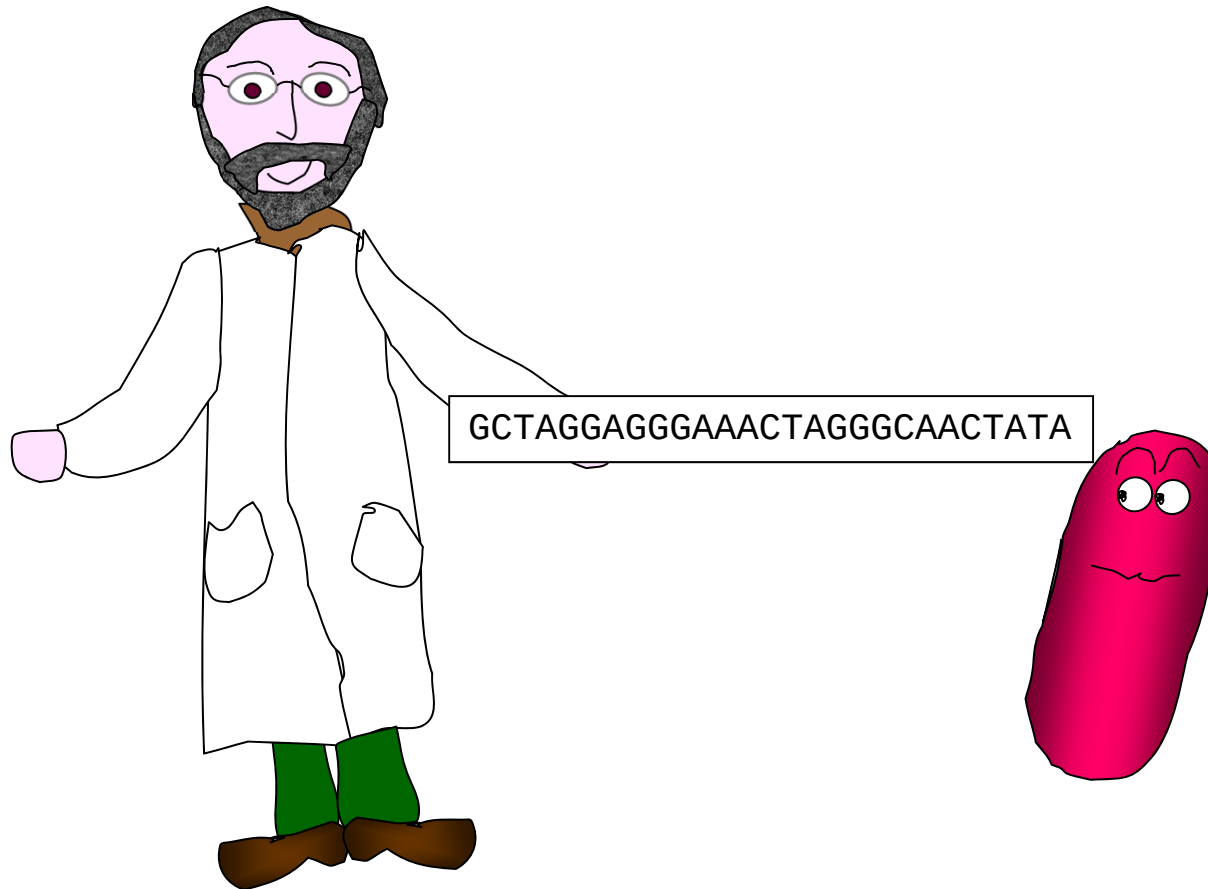


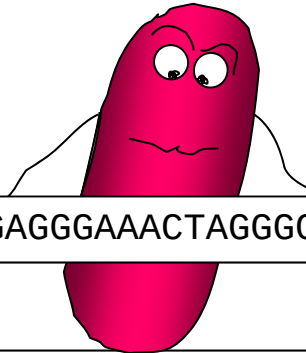
E' anche possibile mutagenizzare il frammento eterologo
per abbassare la quota di codoni rari; specialmente se
questi sono affollati in uno o pochi punti

Mutazioni silenziose causano variazioni di espressione fino a 40x



frammento di DNA eucariotico GCTAGGAGGGGAACTAGGGCAACTATA,
codifica il peptide "A R R E T R Á T I"
Si cerca di farlo esprimere da *E. coli*



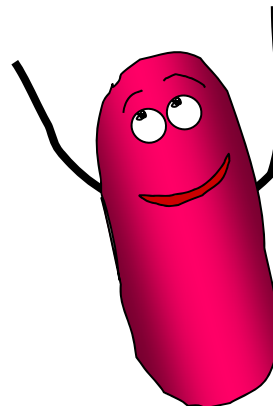


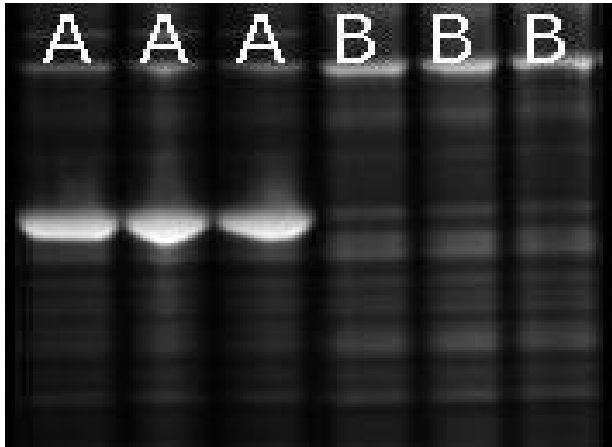
GCTAGGAGGGGAACTAGGGCAACTATA

Secondo l'uso dei codon il risultato sarà: "**ARRÉTRATI**" : l'affollamento di codon rari per Arg rischia di causare uno stallo e un distacco precoce del ribosoma

Per leggere **ARRETRÁTI** bisognerà correggere i codon rari ottenendo un frammento simile a questo

GCT CGT CGT GAG ACT CGT GCA ACT ATA





Con queste procedure molto spesso la resa aumenta

frammento anticorpale con codoni ottimizzati
Stesso frammento in conformazione nativa

ma, a volte questo non succede
PERCHE'?

Welch et al (2009): la resa è influenzata dalla capacità dei tRNA di mantenere lo stato di amino-acilazione e non solo dalla loro disponibilità

**CODONI NEGATIVI
SOSTITUITI CON:**

**tRNA PIU'
STABILI**

**tRNA PIU'
FREQUENTI**

L'espressione migliora più nel primo caso che nel secondo

E sembra legata soprattutto a codoni per Serina, Leucina e Treonina che sono in realtà poco frequenti ma restano stabili molto a lungo

CORPI DI INCLUSIONE

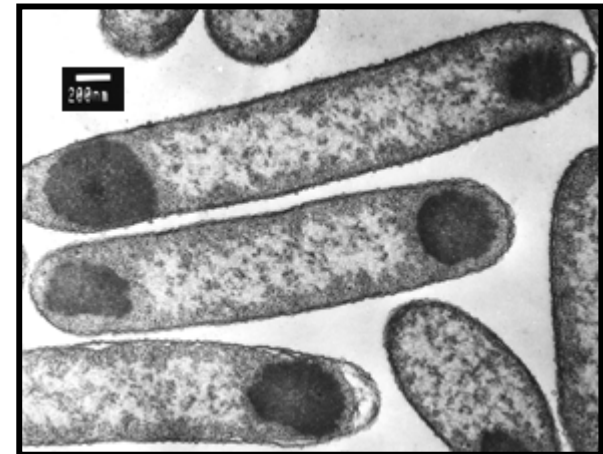
Formazioni opache di $\sim 1 \mu\text{m}$ di diametro,

Insolubili, formati da precipitati di
proteine non conformate correttamente

Vanno trattati in vitro per renaturare
e solubilizzare le proteine, ma i
risultati non sono certi

Strategie possibili:

- ✓ Abbassare la temperatura di incubazione
- ✓ Secernere le proteine nel periplasma
- ✓ Usare ospiti ingegnerizzati
- ✓ Creare proteine di fusione



STRESS METABOLICO

Problema complesso, correlabile a:

- ✓ Scarsità di ossigeno
- ✓ Aumento eccessivo del numero di copie del plasmide
- ✓ Eccessiva quantità di proteine eterologhe accumulate nella cellula

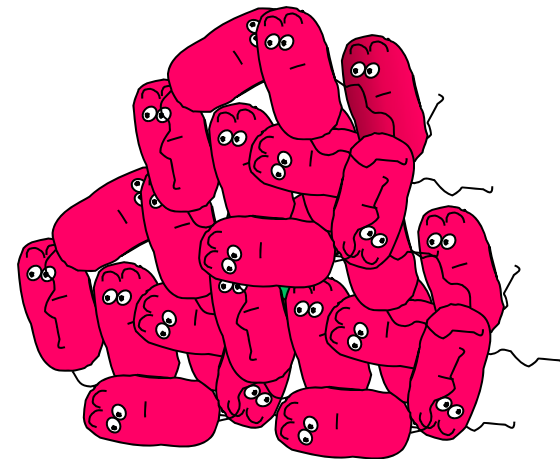
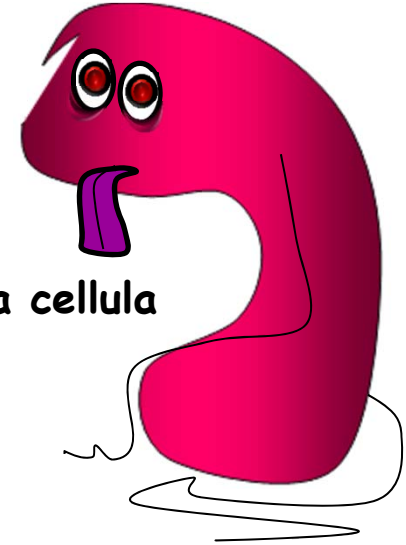
Possibili conseguenze:

- ✓ Perdita del plasmide o del gene eterologo
- ✓ Attivazione delle proteasi
- ✓ Deplezione di aminoacidi e aminoacil-tRNA
- ✓ Errori di traduzione

modificazioni dei polisaccaridi superficiali

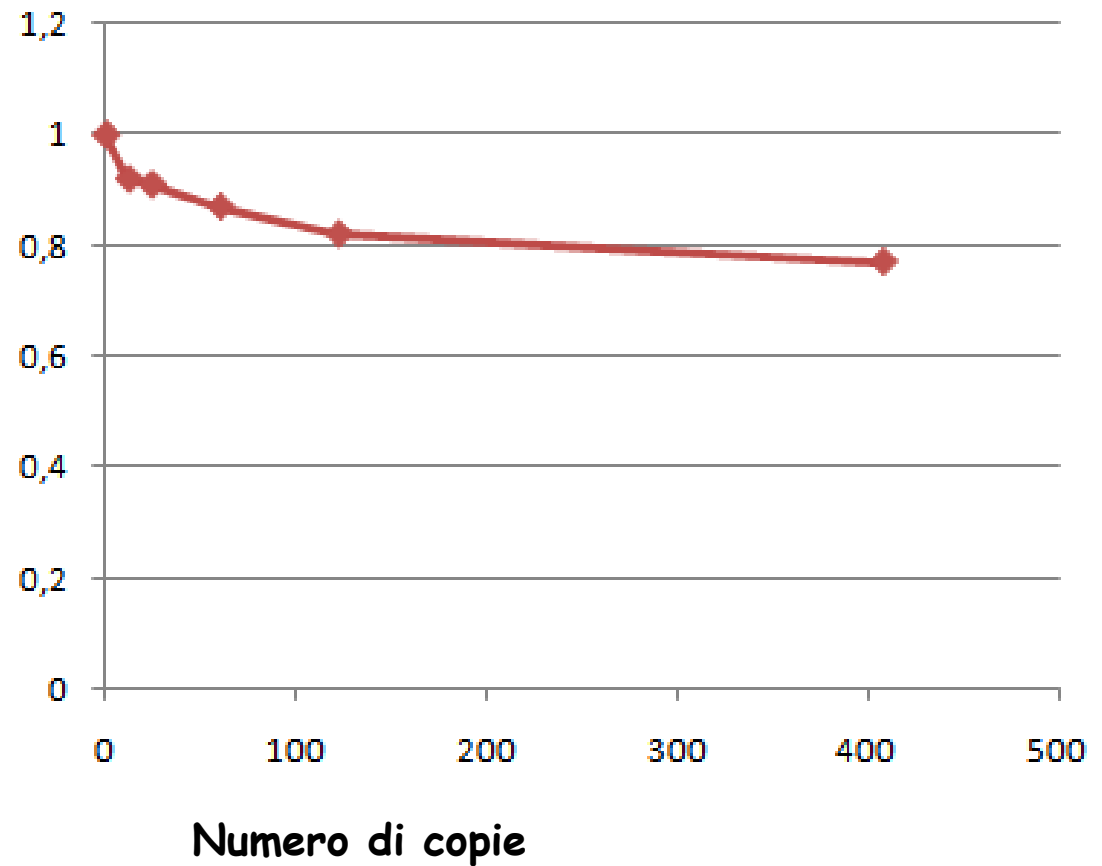
→ aggregazione dei batteri

→ Distribuzione disomogenea di ossigeno e nutrienti a livello delle cellule



Aumentando il numero di copie del plasmide, infine, la velocità di crescita cala, limitando la resa

Tasso di crescita relativo



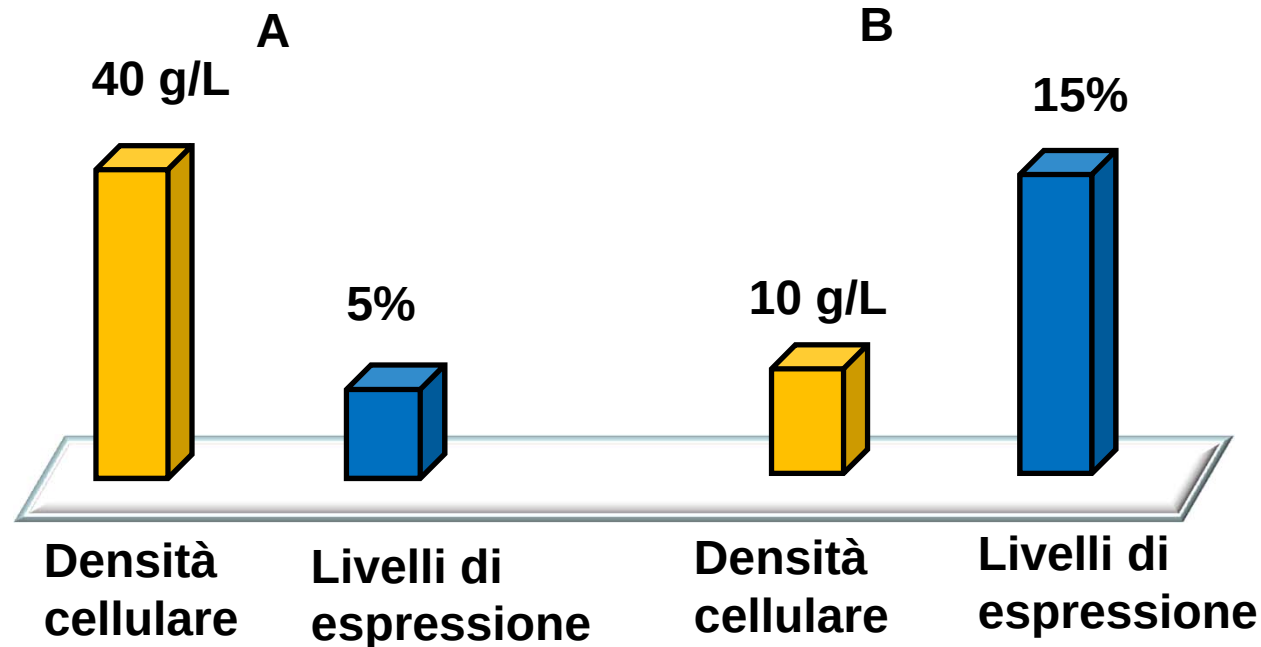
possibili soluzioni:

plasmidi a basso numero di copie

integrazioni nel cromosoma

promotori strettamente regolabili

Diminuire il livello di espressione bilanciando la perdita in resa con l'aumento di biomassa



Rischio: accumulo di acetato → rallentamento della sintesi proteica e del ritmo di replicazione

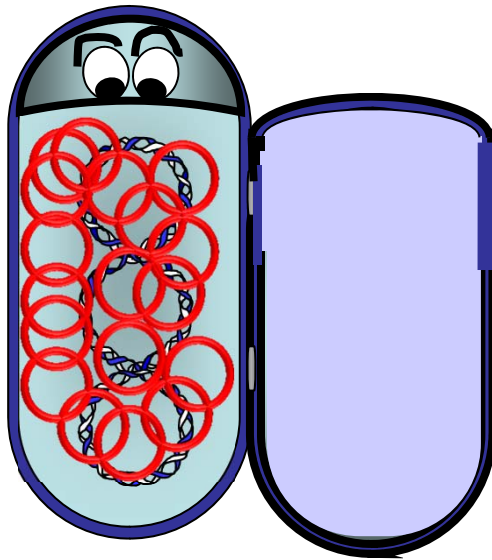
LA MAGGIOR PARTE DEI PROBLEMI DI ESPRESSIONE IN *E. coli* PUO' ESSERE RISOLTA AGENDO SULLE CARATTERISTICHE DEL VETTORE E/O DEL CEPPO OSPITE

I PLASMIDI UTILIZZABILI CON E. COLI SONO MOLTISSIMI
E HANNO CARATTERISTICHE DIVERSE

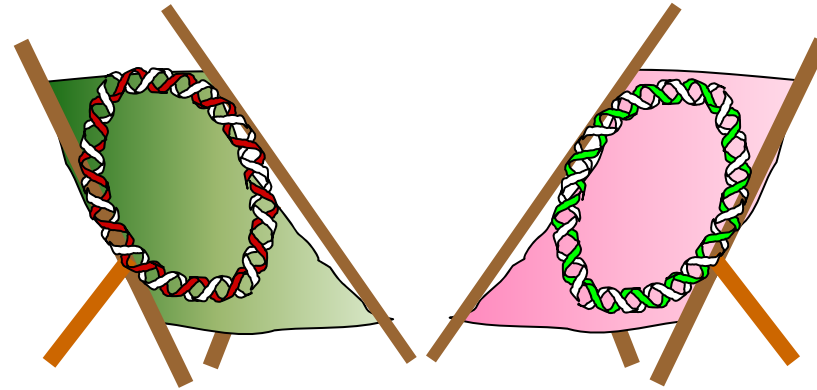
ORIGINE DI REPLICAZIONE

ColE1 $\neq$ pMB1: 1 base

ColE1/pMB1 sono plasmidi rilassati:
dipendono dalle polimerasi dell'ospite



Più frequente ColE1/pMB1



Bloccando la sintesi proteica il
plasmide continua a replicarsi
(polimerasi = enzimi stabili)
spostando il rapporto
plasmide/detriti cellulari

ALTRI REPLICONI

Replicone p15

Diverso gruppo di incompatibilità



**I plasmidi con questa origine
possono coesistere con ColE1/pMB1**

pSC101, plasmide naturale di *Salmonella*

**Replicone stringente a numero
di copie molto basso**

Compatibile con p15 e ColE1/pMB1

Origini F e P1

Usate per cromosomi artificiali, 1-2 copie per cellula

MARCATORI DI SELEZIONE

ANTIBIOTICI

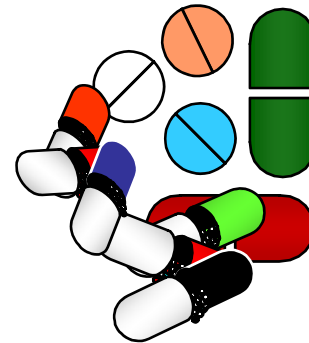
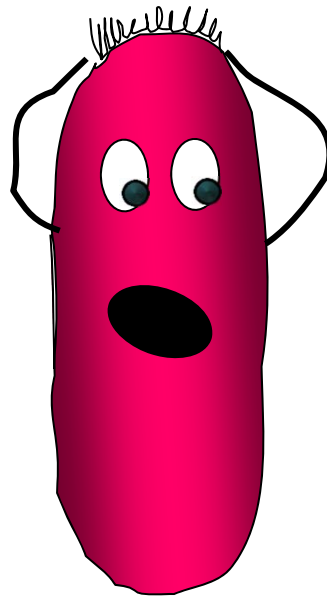
AMPICILLINA

CLORAMFENICOLO

Escherichia coli è molto sensibile

KANAMICINA

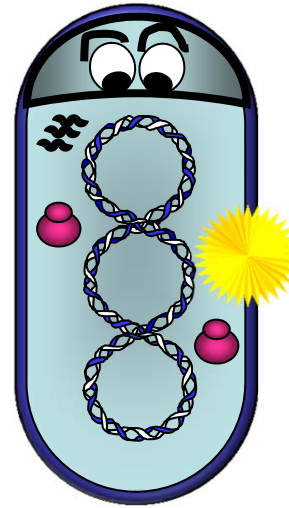
TETRACICLINA



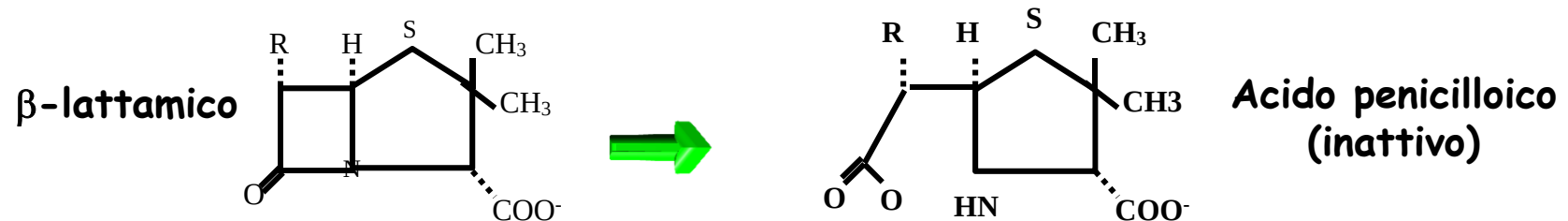
AMPICILLINA

antibiotico β -lattamico, inibisce la sintesi del peptidoglicano

il determinante di resistenza usato sui plasmidi (b/a_{TEM}) codifica una β -lattamasi periplasmica



Attiva sulle penicilline, idrolizza l'anello β -lattamico



AMPICILLINA -vantaggi-

- ✓ facilità d'uso
- ✓ bassi costi
- ✓ rapidità di crescita delle colonie resistenti

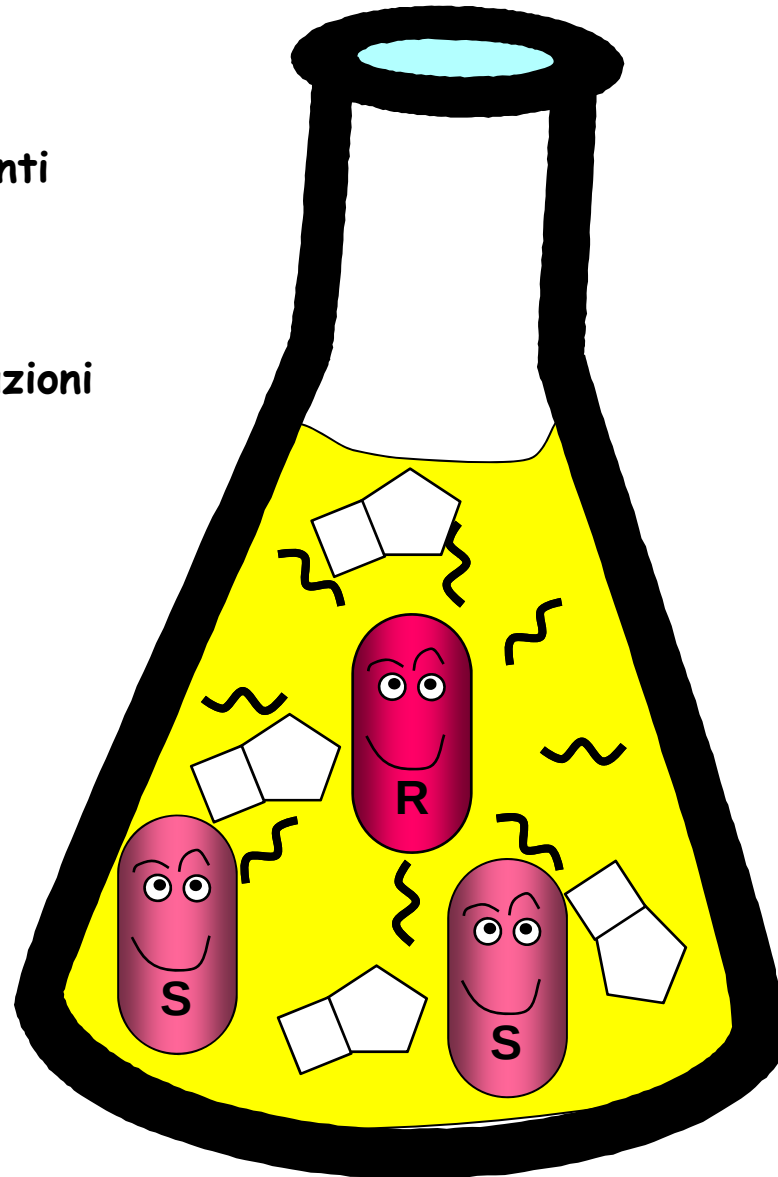
Ottima scelta per il laboratorio

risvolti negativi, scoraggianti per le produzioni

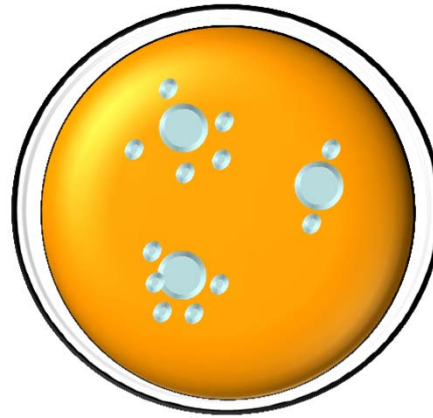
Le cellule Amp^r secernono l'enzima

la concentrazione di antibiotico si riduce

le cellule prive di plasmide
possono crescere



sui terreni solidi
compaiono colonie satelliti

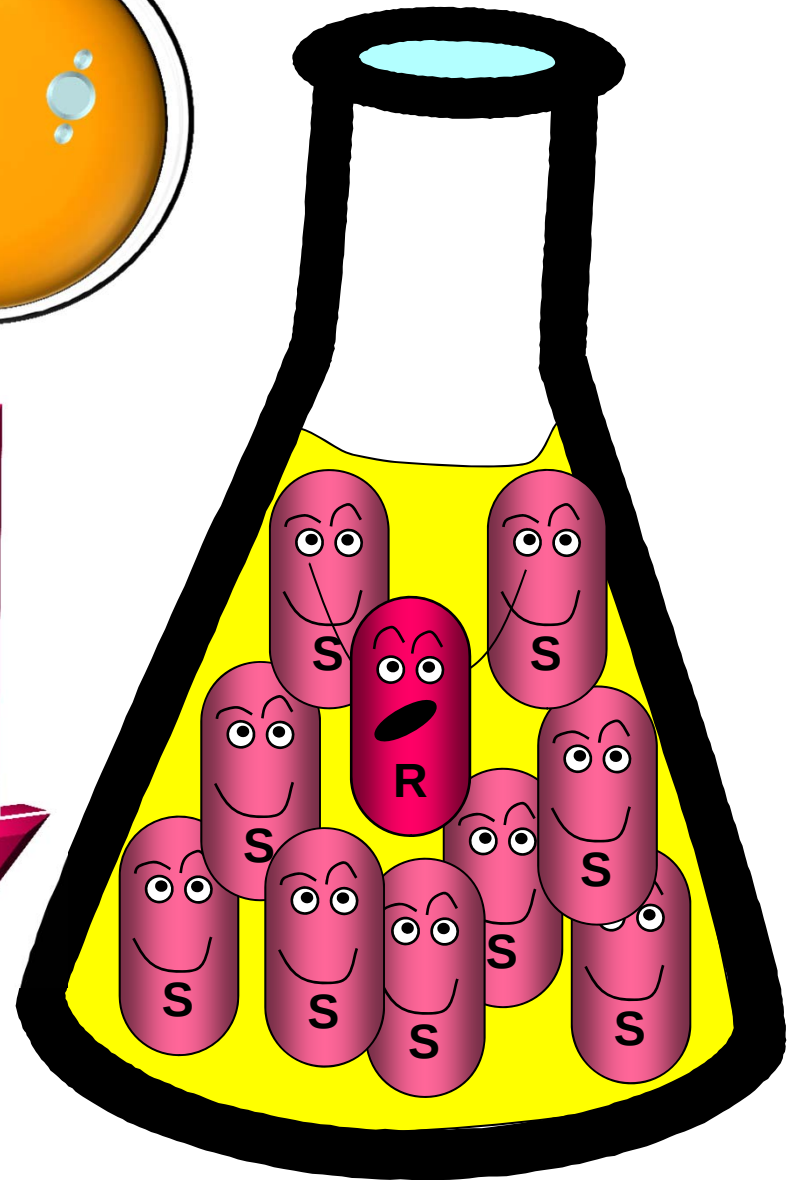


In liquido la perdita di
plasmide diventa imponente

L'ampicillina si degrada
spontaneamente a pH acido

la carbenicillina, più stabile,
può essere un'alternativa

se la selezione è Amp, le colture starter
non dovrebbero crescere più di 8-10 h



CLORAMFENICOLO

Blocca la sintesi proteica legandosi alla subunità 50S del ribosoma

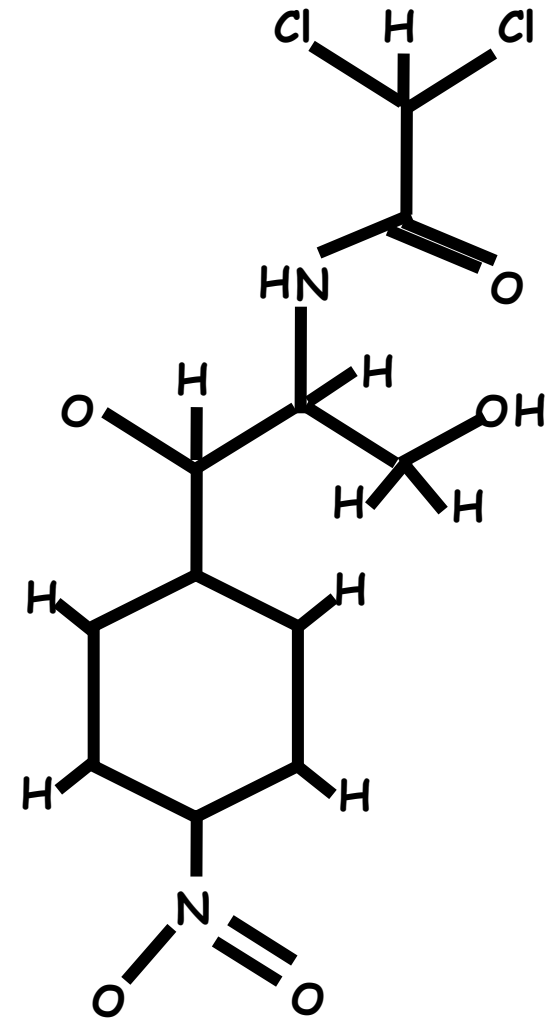
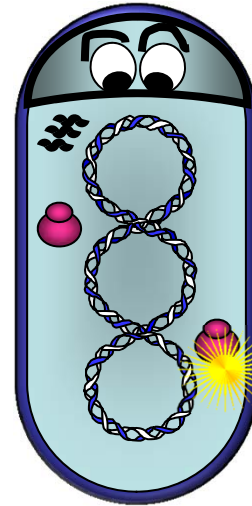
Il determinante usato come marcatore (*cat*) deriva dal trasposone Tn9

proteina citosolica, tetrameric

Cat + CM + acetil CoA



idrossil-acetossi derivati,
incapaci di legare al bersaglio

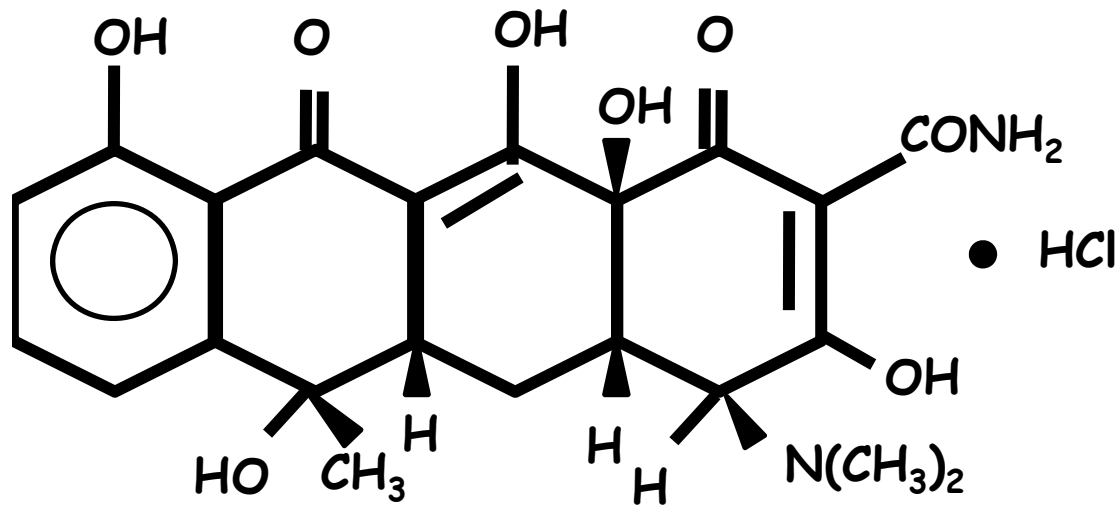


Aspetti negativi: rallenta un po' la crescita, è tossico e potenzialmente carcinogeno, non può essere usato nelle produzioni per uso umano

TETRACICLINE

Composti tetracicclici
batteriostatici prodotti da
Streptomyces o semisintetici

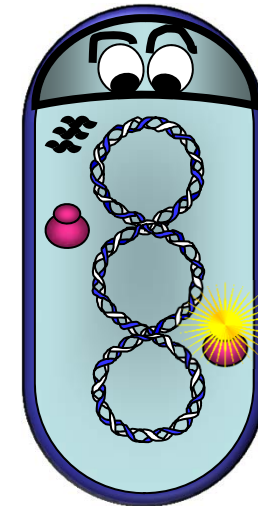
per la selezione si usa l'idrocloruro:
è sufficiente una quantità esigua



Contraggono un legame stabile con la subunità
30S del ribosoma e deformano il sito "A"

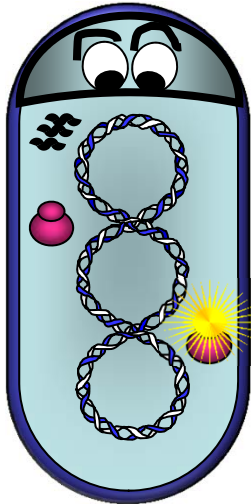
Il gene usato per la selezione è *tetC* da pSC101
ma, sui vettori shuttle se ne usano anche altri

Svantaggi: solubilità limitata (0,4 mg/mL in
acqua; 20 mg/mL in alcol) e fotolabilità



KANAMICINA

antibiotico aminoglicosidico, lega la subunità 30S, bloccando il complesso di inizio



Impedisce il legame con la subunità 50S e l'assemblaggio del ribosoma

Il determinante presente sui plasmidi è derivato dal trasposone Tn903 e conferisce la resistenza a kanamicina e neomicina

