

Metodi per la produzione di proteine ricombinanti

- Lo sviluppo di tecnologie per la produzione di proteine ricombinanti ha dato un impulso enorme alla ricerca sulle proprietà strutturali e funzionali delle proteine
- La produzione di proteine ricombinanti ha ricadute industriali e farmaceutiche importanti

Processi industriali che utilizzano proteine



Agriculture



ANIMAL FEED



Baking



BREWING



DETERGENTS



Fermented Products



FOOD



Fruit Juices



PHARMACEUTICALS



Pulp and Paper



Starch Processing



Textiles

Alcune proteine umane ricombinanti da tempo approvate per uso medico

Insulin (1987)

Interleukin 2 (1989)

Human growth hormone (1991)

Glucagon (1992)

Interferon gamma (1992)

Erythropoietin (1992)

Factor VIII (1993)

Human DNase (1994)

Follitropin (1995)

Calcitonin (1997)

Recombinant Vaccines:

Pertussis

Hepatitis B

Difteria

Lyme disease

Produzione di proteine utilizzabili come farmaci

- Svariate centinaia di proteine umane ad uso terapeutico sono state prodotte tramite la tecnologia rDNA
- Il loro valore commerciale è enorme
 - Si stima che il loro mercato valga oltre 150 miliardi di \$
 - Ormone crescita Genentech 250 milioni \$
 - Insulina Eli Lilly 277 milioni \$
 - Eritropoietina Amgen 2.15 miliardi \$

Proteine umane ricombinanti approvate dalla FDA

Proteina	Industrie	Malattia
DNase I	Genentech	Fibrosi cistica
Eritropoietina	Amgen	Anemia
Ormone crescita	Genentech	Deficienza di GH
Insulina	Eli Lilly	Diabete
IFN- α 2a	Hoffmann-La Roche	Leucemia
Interleukin-2	Chiron	Carcinoma renale
t-PA	Genentech	Infarto

E altre....

Proteine Ricombinanti

Ci sono differenze tra le proteine prodotte per usi industriali e quelle prodotte per uso terapeutico

Queste riguardano la

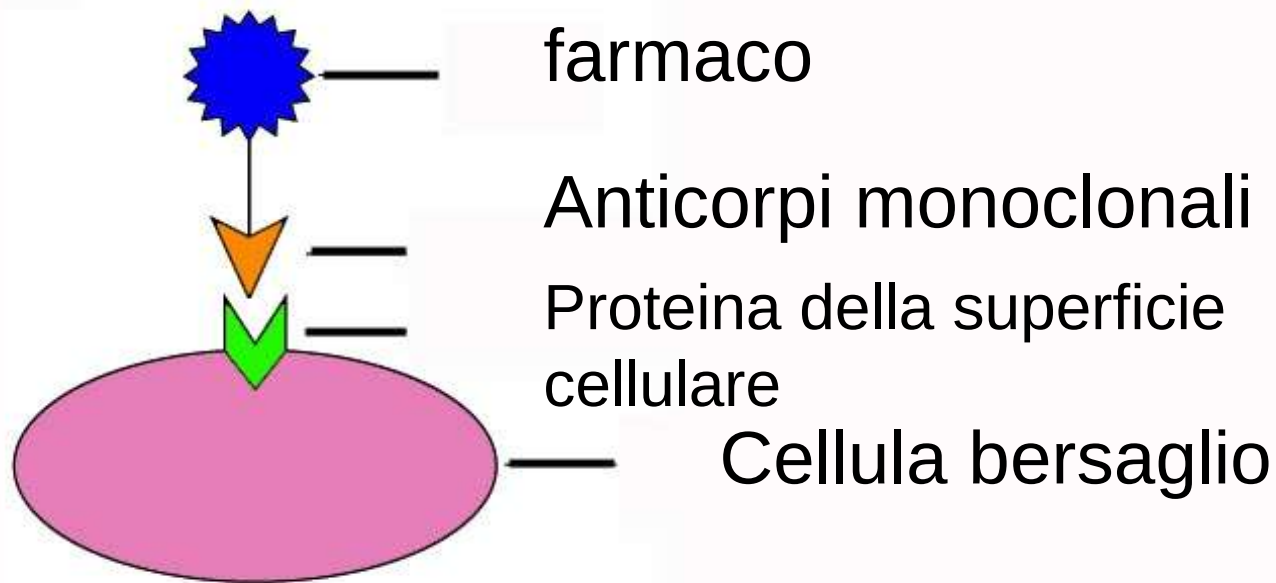
SICUREZZA (identità con il prodotto atteso, assenza di contaminazioni, riproducibilità della strategia di produzione) e i

COSTI DI PRODUZIONE (devono essere bassi per PI, ma possono essere molto alti per PT)

Alcuni obiettivi di interesse biotecnologico

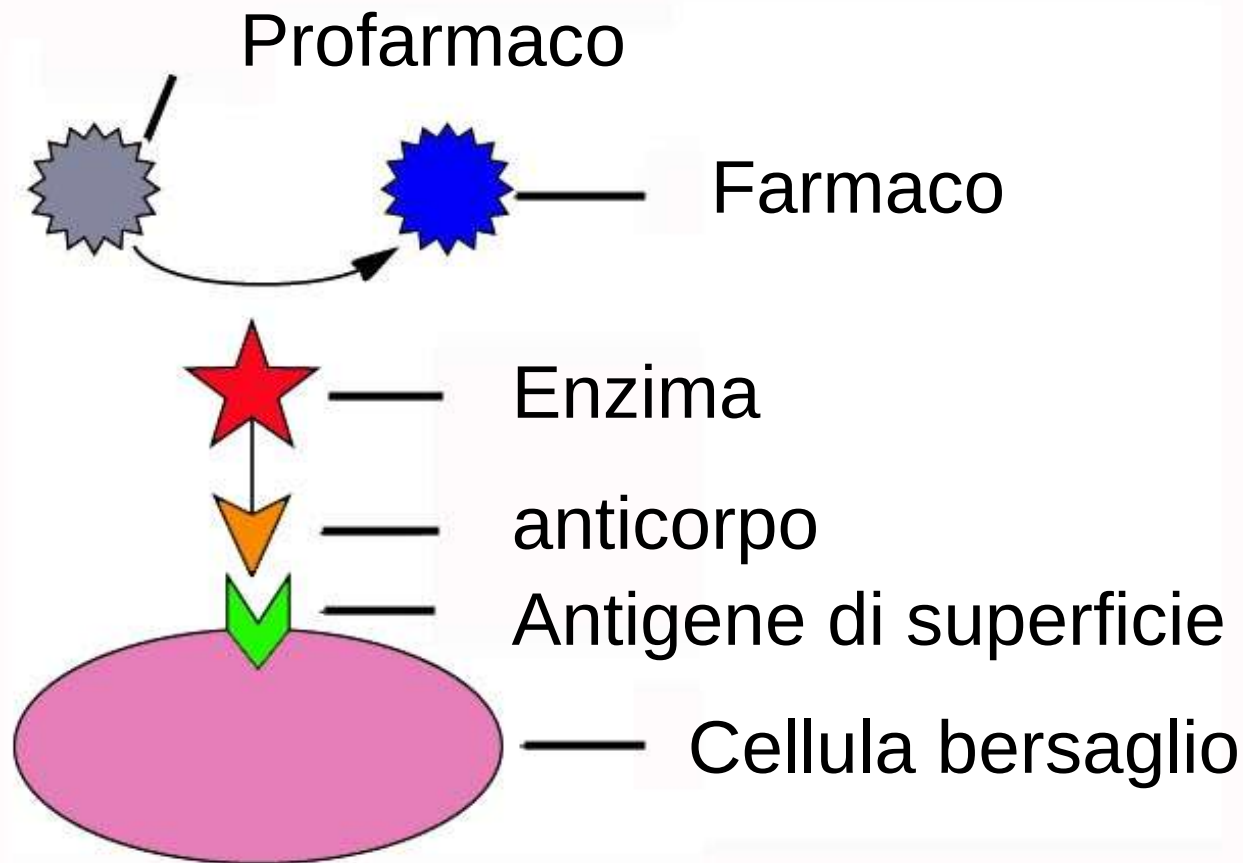
- Riuscire a produrre nuove proteine di interesse medico o industriale
- Massimizzare la produzione di proteine ricombinanti (un'aumentata efficienza di produzione aumenta i profitti)
- Facilitare i processi di purificazione delle proteine
- Costruire proteine diverse da quelle presenti in natura, dotate di nuove e vantaggiose proprietà

Esempio: Veicolazione di farmaci tramite rMAb



La cellula fagocita il complesso proteina-anticorpo

Veicolazione di farmaci tramite rMAb



Strategie per la produzione di proteine ricombinanti

Qual è l'obiettivo ?

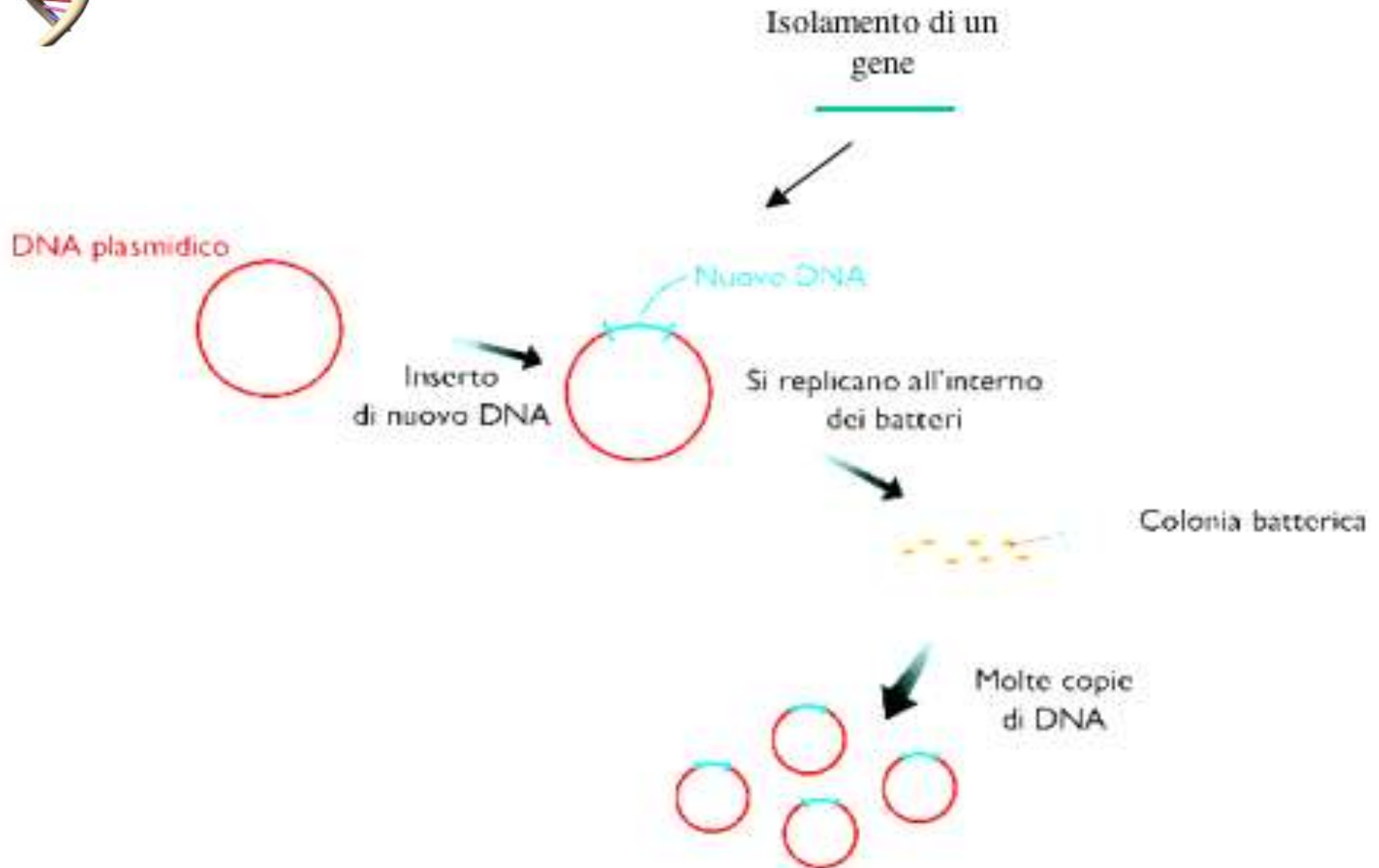
Come scegliere l'ospite in cui produrre la proteina?

Quanta proteina serve?

Come esprimere la proteina (espressione costitutiva, inducibile..)?



Clonaggio molecolare



Some Enzymes Used in Recombinant DNA Technology

Enzyme(s)	Function
Type II restriction endonucleases	Cleave DNAs at specific base sequences
DNA ligase	Joins two DNA molecules or fragments
DNA polymerase I (<i>E. coli</i>)	Fills gaps in duplexes by stepwise addition of nucleotides to 3' ends
Reverse transcriptase	Makes a DNA copy of an RNA molecule
Polynucleotide kinase	Adds a phosphate to the 5'-OH end of a polynucleotide to label it or permit ligation
Terminal transferase	Adds homopolymer tails to the 3'-OH ends of a linear duplex
Exonuclease III	Removes nucleotide residues from the 3' ends of a DNA strand
Bacteriophage λ exonuclease	Removes nucleotides from the 5' ends of a duplex to expose single-stranded 3' ends
Alkaline phosphatase	Removes terminal phosphates from either the 5' or 3' end (or both)



Esempi di enzimi di restrizione

Microrganismo	Enzima di restrizione	Sequenza bersaglio
<i>Anabaena variabilis</i>	<i>Ava</i> I	↓C↓PyCGPuG
<i>Bacillus amyloliquefaciens</i> H	<i>Bam</i> HI	G↓GATCC
<i>Bacillus albinum</i>	<i>Bgl</i> II	A↓GATCT
<i>Brevibacterium albinum</i>	<i>Ba</i> I	TGGCCA↓
<i>Escherichia coli</i> RY13	<i>Eco</i> RI	G↓AATTC
<i>Escherichia coli</i> R245	<i>Eco</i> RII	↓CCAGG
<i>Haemophilus aegyptius</i>	<i>Hae</i> II	PuGCGC↓Py
<i>Aemophilus aegyptius</i>	<i>Hae</i> III	G↓GCC
<i>Haemophilus haemophilus</i>	<i>Hha</i> I	G↓CGC
<i>Haemophilus influenzae</i> RD	<i>Hind</i> II	GTPy↓PuAC
<i>Haemophilus influenzae</i> RD	<i>Hind</i> III	A↓AGCTT
<i>Haemophilus parainfluenzae</i>	<i>Hpa</i> I	GTTA↓AC
<i>Haemophilus parainfluenzae</i>	<i>Hpa</i> II	C↓CGG
<i>Klebsiella pneumoniae</i>	<i>Kpn</i> I	GGTACC↓
<i>Moraxella bovis</i>	<i>Mbo</i> I	↓GATC
<i>Providencia stuartii</i>	<i>Pst</i> I	CTGCA↓G
<i>Streptomyces albus</i>	<i>Sa</i> I	G↓TCGAC
<i>Serratia marcescens</i>	<i>Sma</i> I	CCCGGG↓
<i>Xanthomonas malvacearum</i>	<i>Xma</i> I	C↓CCGGG



Vettori di clonaggio

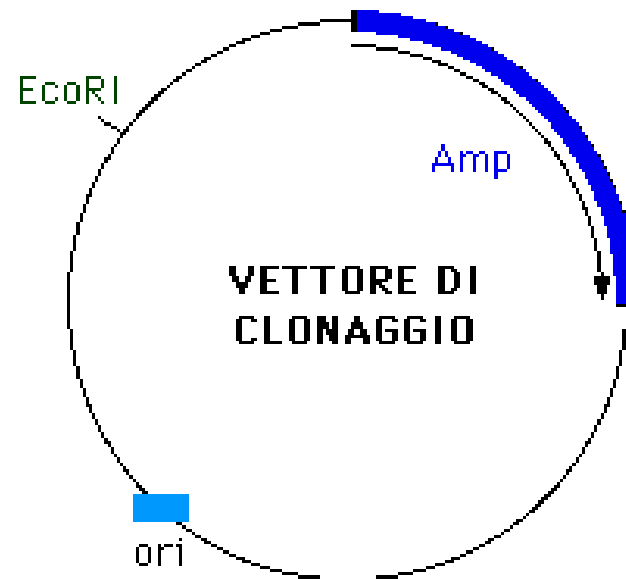
● Plasmidi (da 0,1 a 10-15 Kb)	naturali	Col E1 pSC101	Vettori procariotici
	artificiali	pBR322 pUC8 pEMBL8	
● Fagi (da 8 a 22 kb)	lambda M13		
● Cosmidi (da 32 a 45 kb)			
● Minicromosomi artificiali (es. YAC; da 100 a 2000 kb)			



VETTORI DI CLONAGGIO

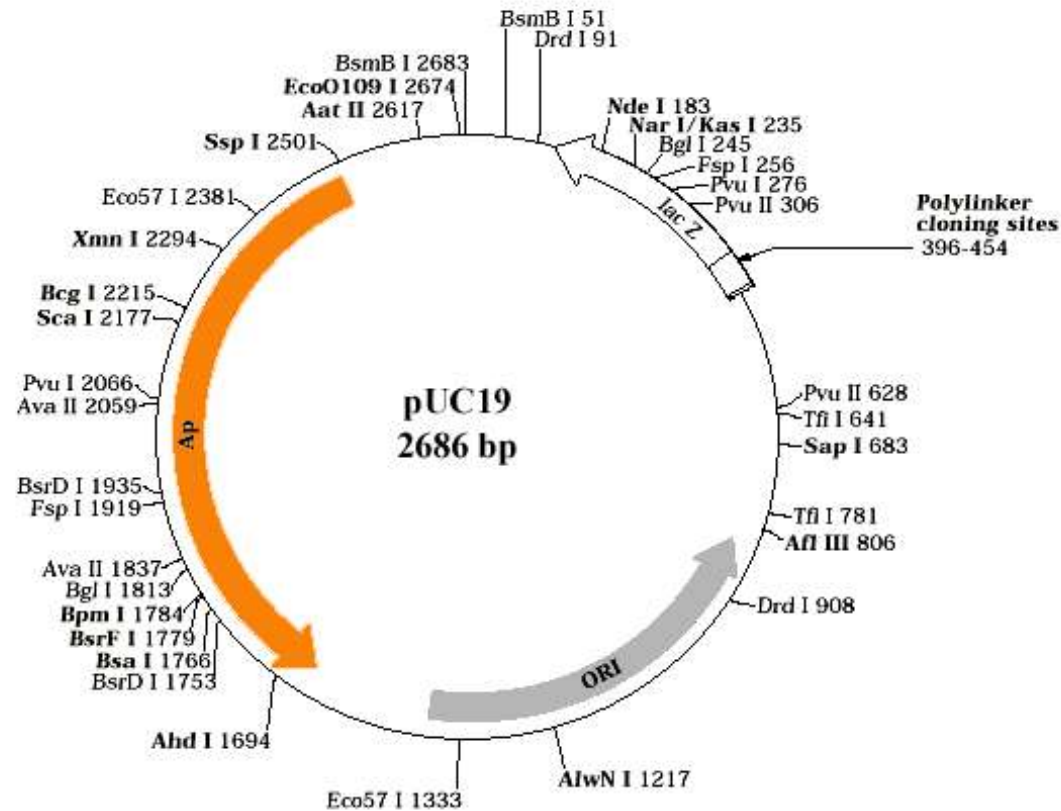
Dai plasmidi batterici naturali sono derivati i vettori di clonaggio, le cui caratteristiche **essenziali** sono:

- Origine di replicazione
- Marcatore selezionabile
- Siti di restrizione unici

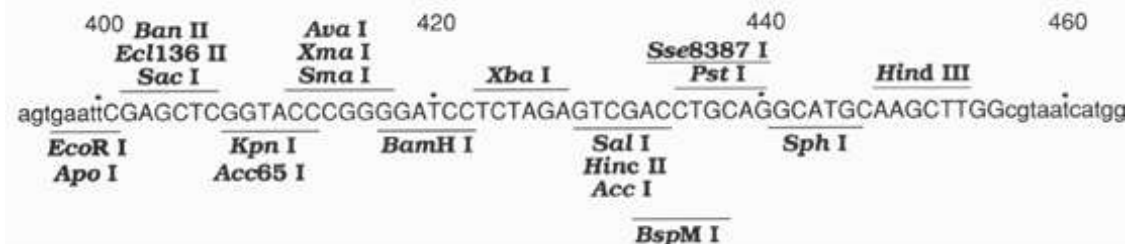


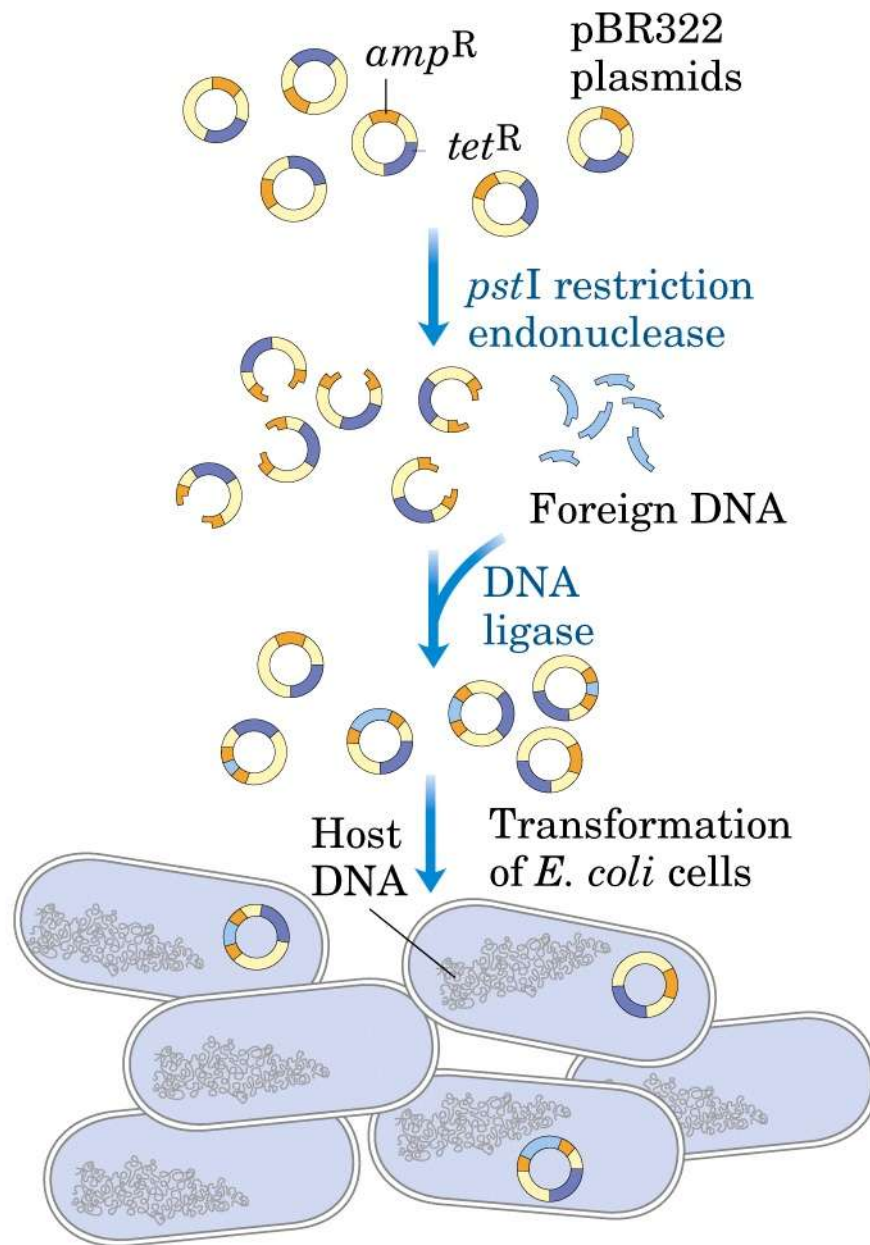


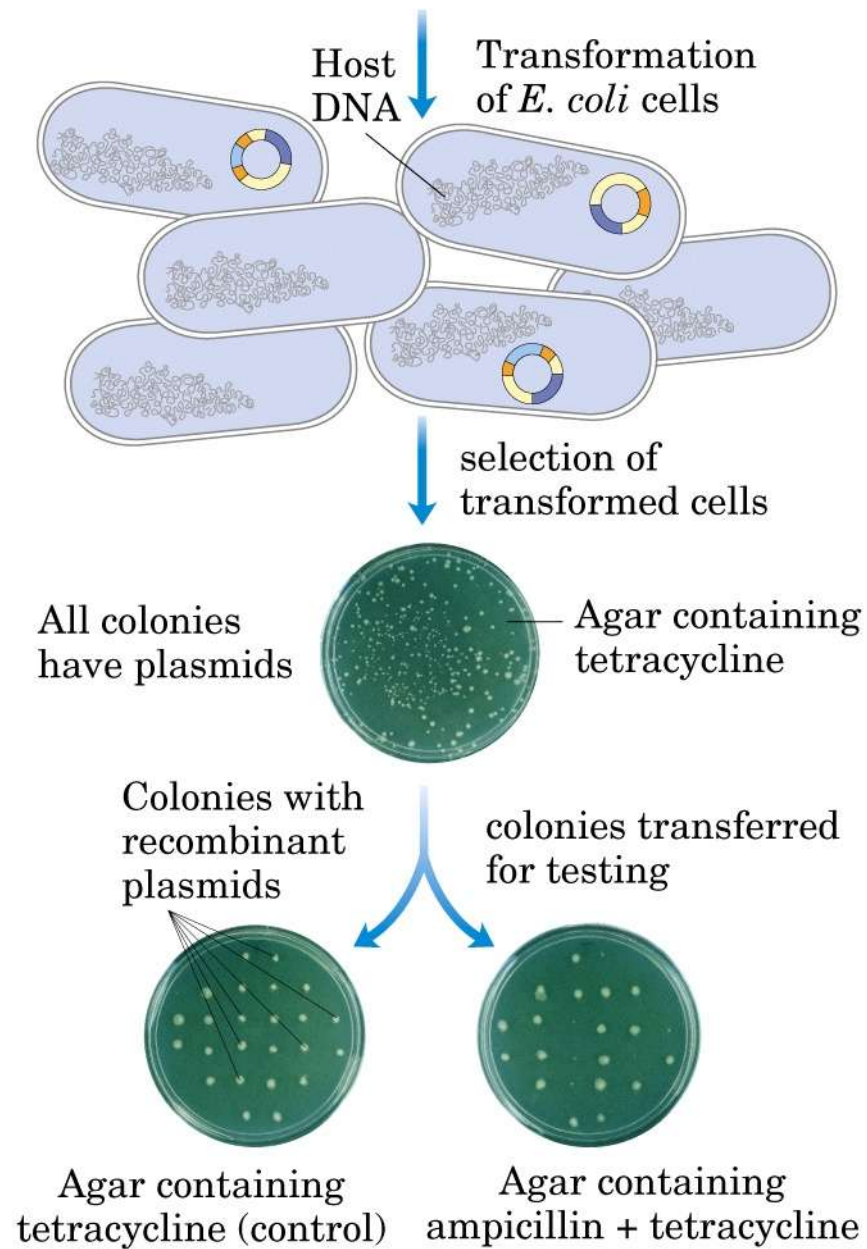
Vettore plasmidico



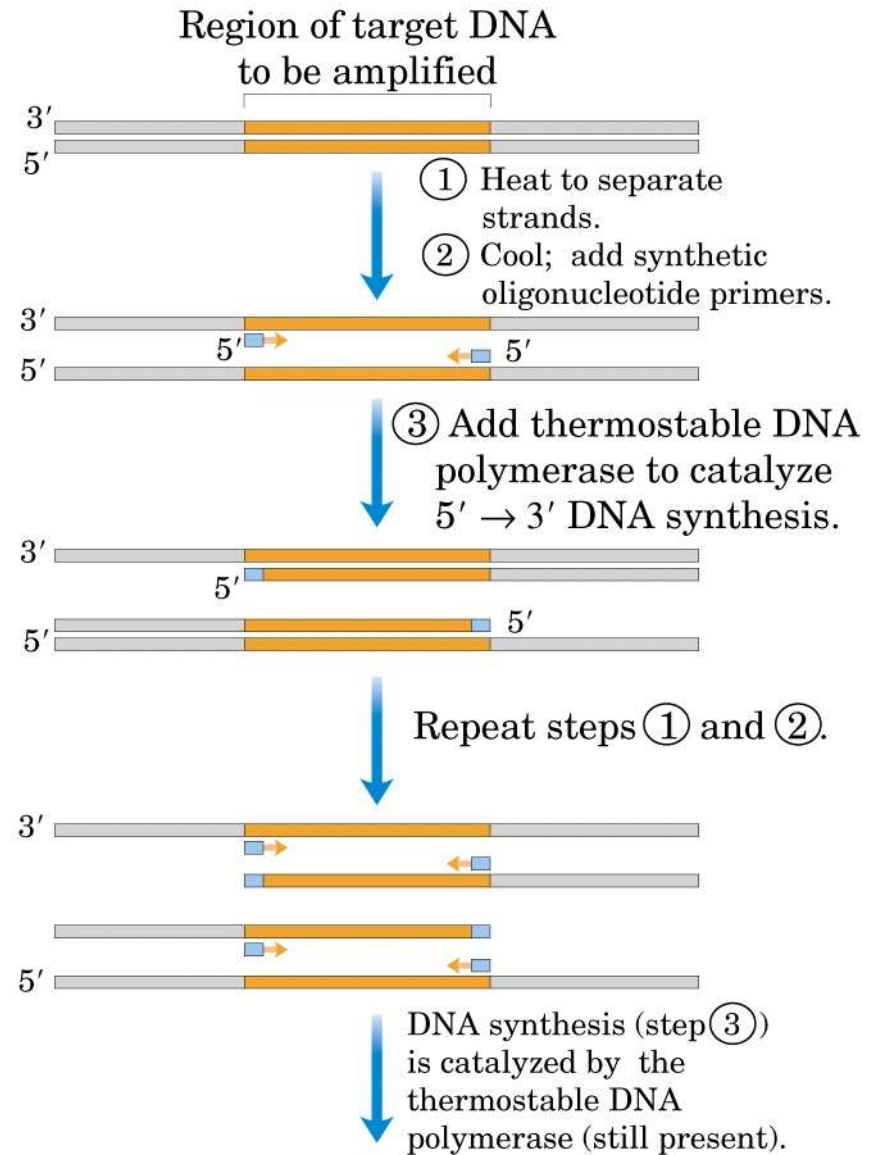
polylinker region



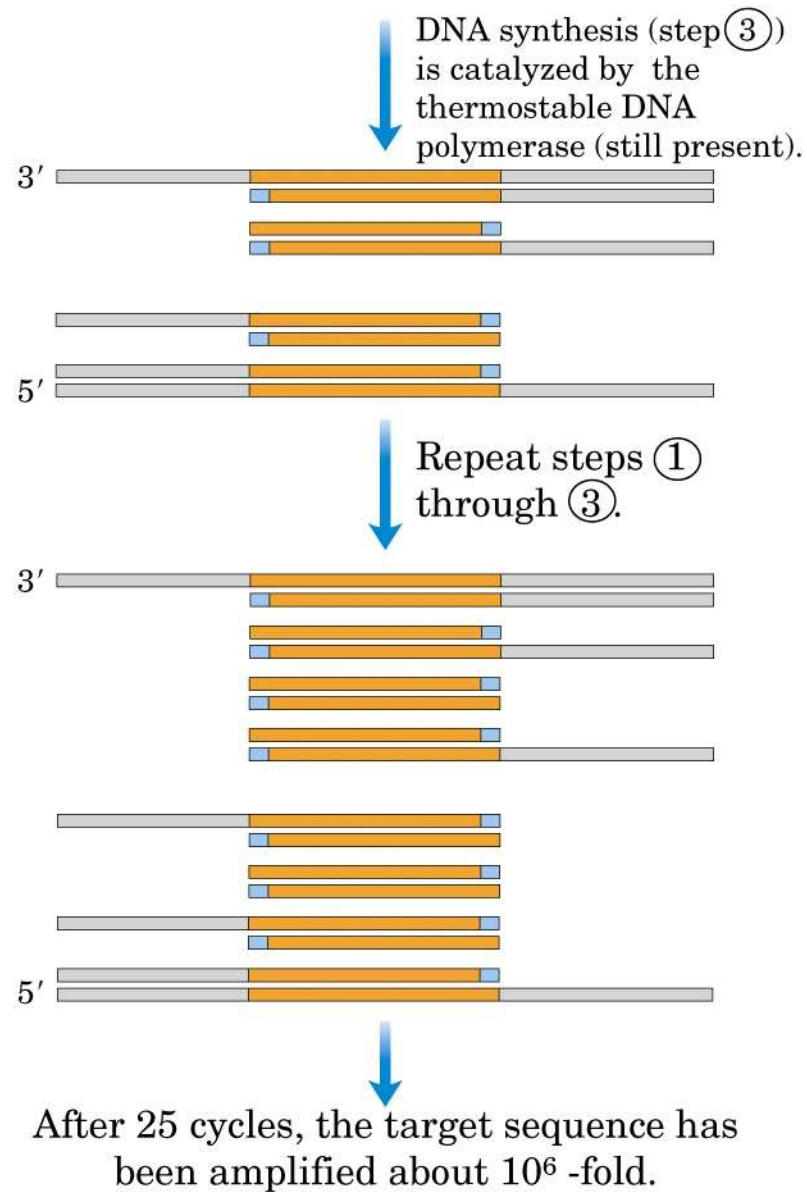




Clonaggio mediante PCR



(a)



(a)

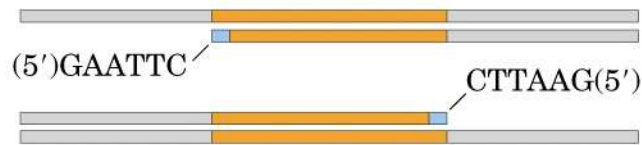


① Heat to separate strands.

② Anneal primers containing noncomplementary regions with cleavage site for restriction endonuclease.



Replication



PCR



*Eco*RI endonuclease



Clone by insertion
at an *Eco*RI site
in a cloning vector.

(b)

Sistemi di espressione per la produzione efficiente di proteine ricombinanti

Batteri (*E. coli*)

Lieviti (*S. cerevisiae* ; *P. pastoris*)

Cellule in coltura (cellule di insetti e mammiferi)

Modelli animali (mucche, pecore) o vegetali

La **scelta** deve tenere conto di diversi **fattori**:

- modificazioni post-traduzionali delle proteine
- problemi a livello di regolazione genica
- costi e possibilità di scaling up

I microrganismi sono facilmente manipolabili, economici, modificabili in tempi rapidi. Tuttavia, le proteine prodotte nei batteri sono prive di modificazioni post-traduzionali

Le cellule di lievito e di insetto consentono di ottenere glicosilazioni, ma queste sono spesso specie-specifiche

Le coltivazione di cellule di mammifero è lenta, costosa e difficilmente consente l'ottenimento di proteine in alta resa

L'uso di animali come bioreattori ha alcuni vantaggi, ma notevoli limitazioni

Microbial expression hosts – *E. coli*

First choice for medium-small proteins

ADVANTAGES

- Numerous specific vectors
- Very high transformation efficiencies
- Genetics well established
- Very fast and cheap growth
- High expression levels
- Very well understood fermentations
- Simple cell recovery and lysis
- Easy scale up procedures

DISADVANTAGES

- Not naturally competent
- Low expression yields (sometimes)
- No post-translational modification of rProtein (glycosylations, acetylation, phosphorylation)
- Often proteins are denatured (need refolding)
- Problems with multimeric proteins and S-S bridges
- proteolysis
- rProtein not exported

Microbial expression hosts – *Saccharomyces cerevisiae*

ADVANTAGES

- Moderate/high expression yields
- rProtein glycosylated
- Cheap and fast growth
- rProtein exported
- Very well understood fermentations
- Easy scale-up
- Simple cell recovery

DISADVANTAGES

- Low-medium transformation efficiency
- Limited range of vector systems
- rGenes must be stably incorporated
- Limited post-translational modifications
- Cell lysis

Microbial expression hosts – *Pichia pastoris* and *Kluyveromyces sp.*

ADVANTAGES

- Moderate expression yields
- rProtein glycosylated
- rProtein exported
- Simple cell recovery

DISADVANTAGES

- Low-medium transformation efficiencies
- Difficult fermentations
- Limited range of vector systems
- rGenes must be stably incorporated
- Unusual codon usages

Alternative microbial expression hosts – *Bacillus sp.*

ADVANTAGES

- Very high expression yields (>10 g/L)
- Simple fermentations
- rProtein exported
- Simple cell recovery
- Naturally competent

DISADVANTAGES

- Limited range of vector systems
- Optimised vector systems are proprietary
- No glycosylation
- Variable transformation efficiencies

Insect cell expression

Specific host-vector system; Sf9 cultured insect cells and *Baculovirus* vectors (circular double-stranded genome ranging from 80–180 kbp)

- **Advantages**

- Higher eukaryote-like glycosylation
- Protein export
- Scalable to large volumes

- **Disadvantages**

- Highly specific transformation system (Baculovirus)
- Quite new expression system
- Fastidious culture
- Very slow doubling time (20h)
- More expensive than microorganisms
- Not so easy to scale up

Animal cell culture

- **Advantages**

- Exact glycosylation
- Can be transformed to continuous cell lines
- Some (e.g., CHO cells) grow in suspension culture
- Some carcinoma-derived cells (e.g., HeLa cells) have rapid growth rates
- Extracellular expression

- **Disadvantages**

- Some grow as adherent layer (anchorage dependent)
- Fastidious growth requirements
- Very slow doubling time (15-25h)
- Very susceptible to contamination
- Very sensitive to shear stress
- Very expensive media
- Low yields

Plant cell culture

- Advantages

- Suspension cultures
- Low aeration required
- rProtein directed to vacuole

- Disadvantages

- Restricted range of transformation systems
 - *Agrobacterium* for dicots
 - Biolistics for monocots
 - Specific plant viruses
- Very slow doubling times (15h – 48h)
- Very sensitive to shear stress
- Grow heterotrophically but require light for optimum growth

Other expression systems

Whole animal expression

- Transformation of unfertilized embryo
- Use of tissue-specific promoters (e.g., lactose-specific promoter for expression of proteins in milk)



- Be careful of post-translational modification identity

Escherichia coli è il sistema di scelta per tutte le proteine di medie dimensioni (100-250 aminoacidi) prive di modificazioni post-traduzionali.

In ogni caso è utile per valutare la funzione delle modificazioni e per la preparazione di antigeni

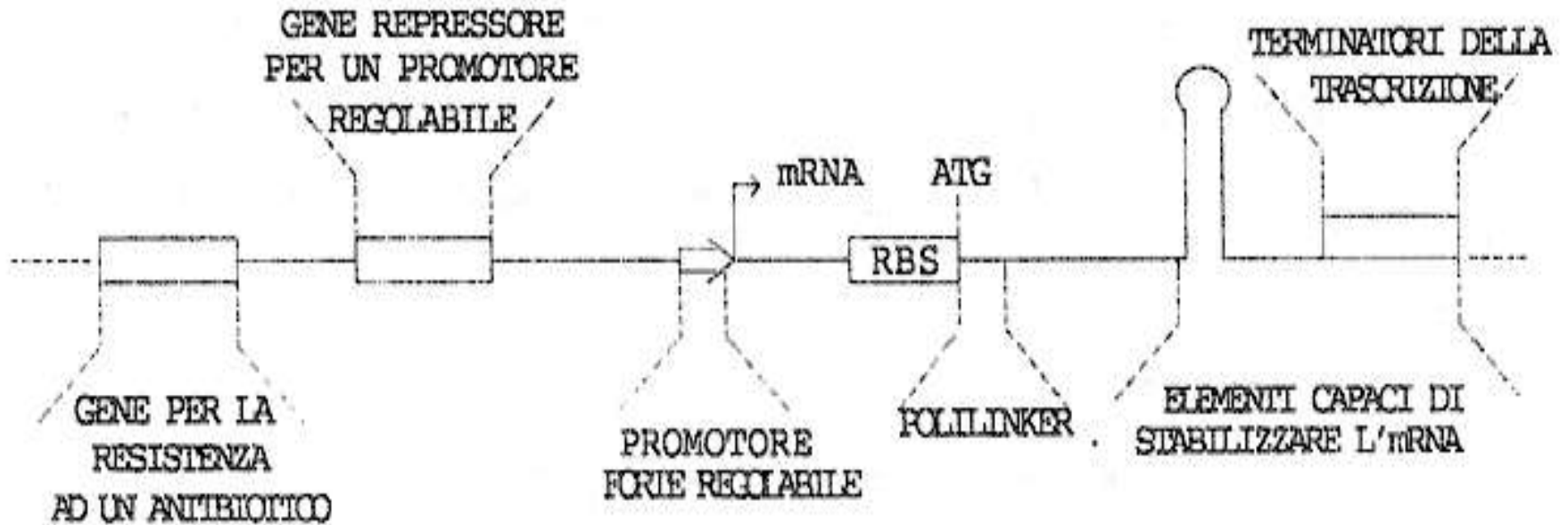
E possibile esprimere la proteina in diversi modi:

- a) Espressione diretta citoplasmatica (proteina solubile nel citoplasma)
- b) Espressione diretta periplasmatica (proteina solubile nel periplasma)
- c) Espressione di una fusione.

Manipolazione dell' espressione genica nei procarioti

- Clonare un gene non è sufficiente per garantire l'espressione ad alti livelli della proteina codificata
- Spesso è necessario introdurre modifiche al vettore d'espressione, al gene stesso o modificare l'ospite.

Caratteristiche di un vettore di espressione



Proprietà modificabili del sistema di espressione

- 1) Promotore
- 2) Sequenza di Shine Dalgarno-ribosome binding site (RBS)
- 3) Efficienza di traduzione
- 4) Localizzazione della proteina (secreta?)
- 5) Numero di copie del plasmide

Plasmide vs. cromosoma

Tante vs. 1

- 6) Organismo ospite
- 7) Problemi legati alla proteina (vedi oltre)

Esprimere proteine in un sistema eterologo non è sempre banale!

- Nessuna strategia funzionerà in tutti i casi
- Ogni proteina è unica
- Può essere necessario apportare più modifiche alle strategie standard

Gene espresso a partire da un promotore forte e regolabile

- La trascrizione è spesso il punto di controllo più importante per la regolazione dell'espressione genica
- La trascrizione è controllata dal promotore
- I promotori devono essere
 - Forti
 - Possibilmente Regolabili

Promotori forti e regolabili

Forte

- Alta affinità per l'RNA polimerasi
- Legame forte - trascritto ad alta frequenza
- Legame debole - RNA Pol si stacca → no transcription

Regolabile

- L'espressione può essere controllata dal ricercatore, tramite induttori e repressore

Alcuni promotori usati frequentemente

- lac
- trp
- tac
- λ p^L
- gene 10 del fago T7

Promotore lac

- Operone lac
- Regolazione negativa
 - Il repressore di lac si lega alla sequenza operatore
 - Induttore si lega al repressore, rendendolo non funzionale
 - Induttore = analogo del lattosio, IPTG

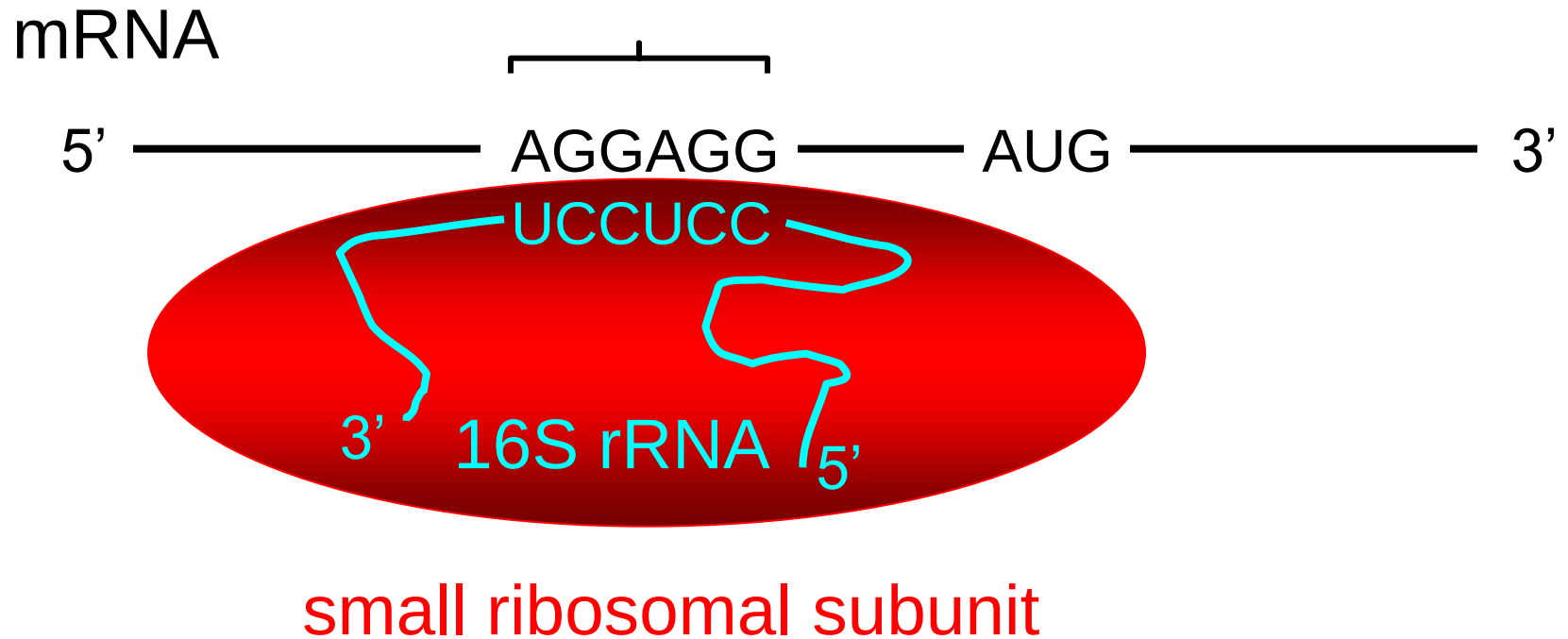
Promotori “Leaky”

- Alcuni promotori (lac, ad esempio) non riescono ad essere controllati in modo molto stringente.
- Si ha quindi sempre un basso livello di espressione.

Migliorare la Traduzione

- I promotori forti producono grandi quantità di mRNA
- E' però necessario che la traduzione del messaggero sia efficiente
- DNA $\longrightarrow$ RNA $\longrightarrow$ Protein

La sequenza di Shine e Dalgarno



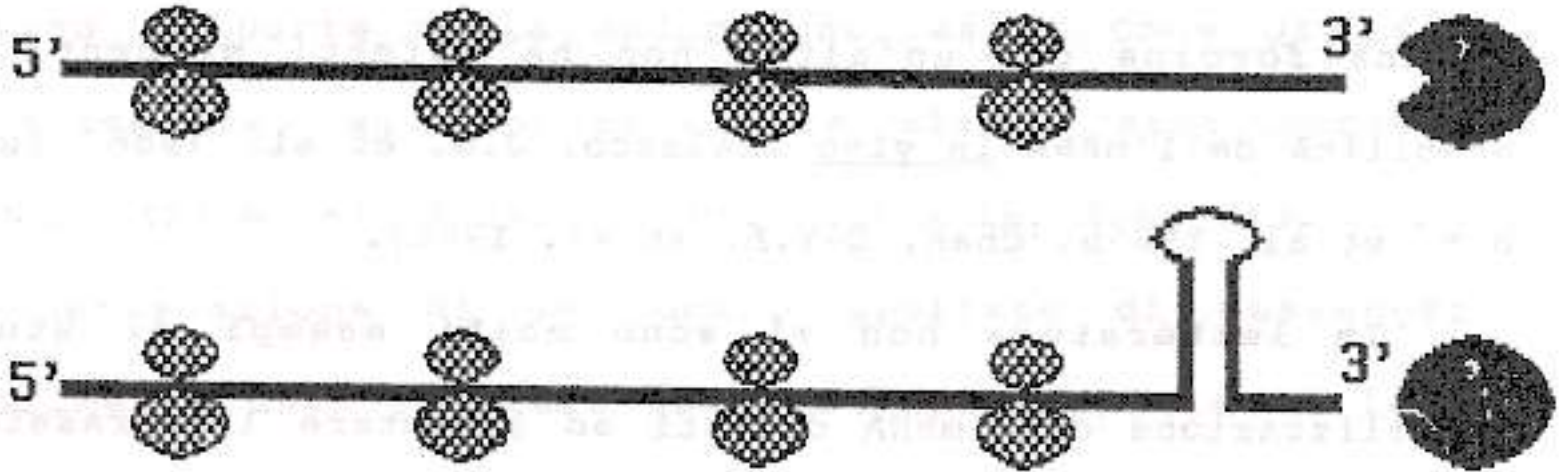
Migliorare la traduzione

- Uso dei codoni
 - Molti codoni per lo stesso aminoacido
 - Alcuni sono usati di rado
 - poco tRNA corrispondente
- Differenti organismi usano il codice in modo diverso
- Obiettivo: ottimizzare i codoni

Table 3.2 The genetic code and codon usage in *E. coli* and humans

Codon	Amino acid	Frequency of use in:	
		<i>E. coli</i>	Humans
GAG	Glutamic acid	0.30	0.59
GAA	Glutamic acid	0.70	0.41
CGG	Arginine	0.08	0.19
CGA	Arginine	0.05	0.10
CGU	Arginine	0.42	0.09
CGC	Arginine	0.37	0.19
AGG	Arginine	0.03	0.22
AGA	Arginine	0.04	0.21
CCG	Proline	0.55	0.11
CCA	Proline	0.20	0.27
CCU	Proline	0.16	0.29
CCC	Proline	0.10	0.33
UGA	Stop	0.30	0.61
UAG	Stop	0.09	0.17
UAA	Stop	0.62	0.22

Stabilizzare il messaggero



Problemi legati alla proteina

- 1) Incapacità della proteina di assumere la corretta struttura 3D nell'ambiente batterico
- 2) Assenza di cofattori essenziali
- 3) Mancata modificazione post-traduzionale
- 4) Degradazione proteolitica
- 5) Tossicità della proteina

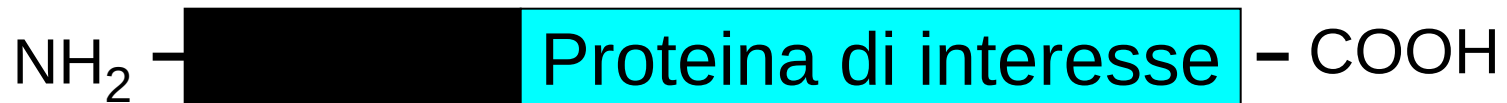
Esporto della Proteina

La stabilità della proteina può aumentare
(meno proteasi, differenze nella
formazione dei legami disolfuro)

La proteina può essere isolata più
facilmente

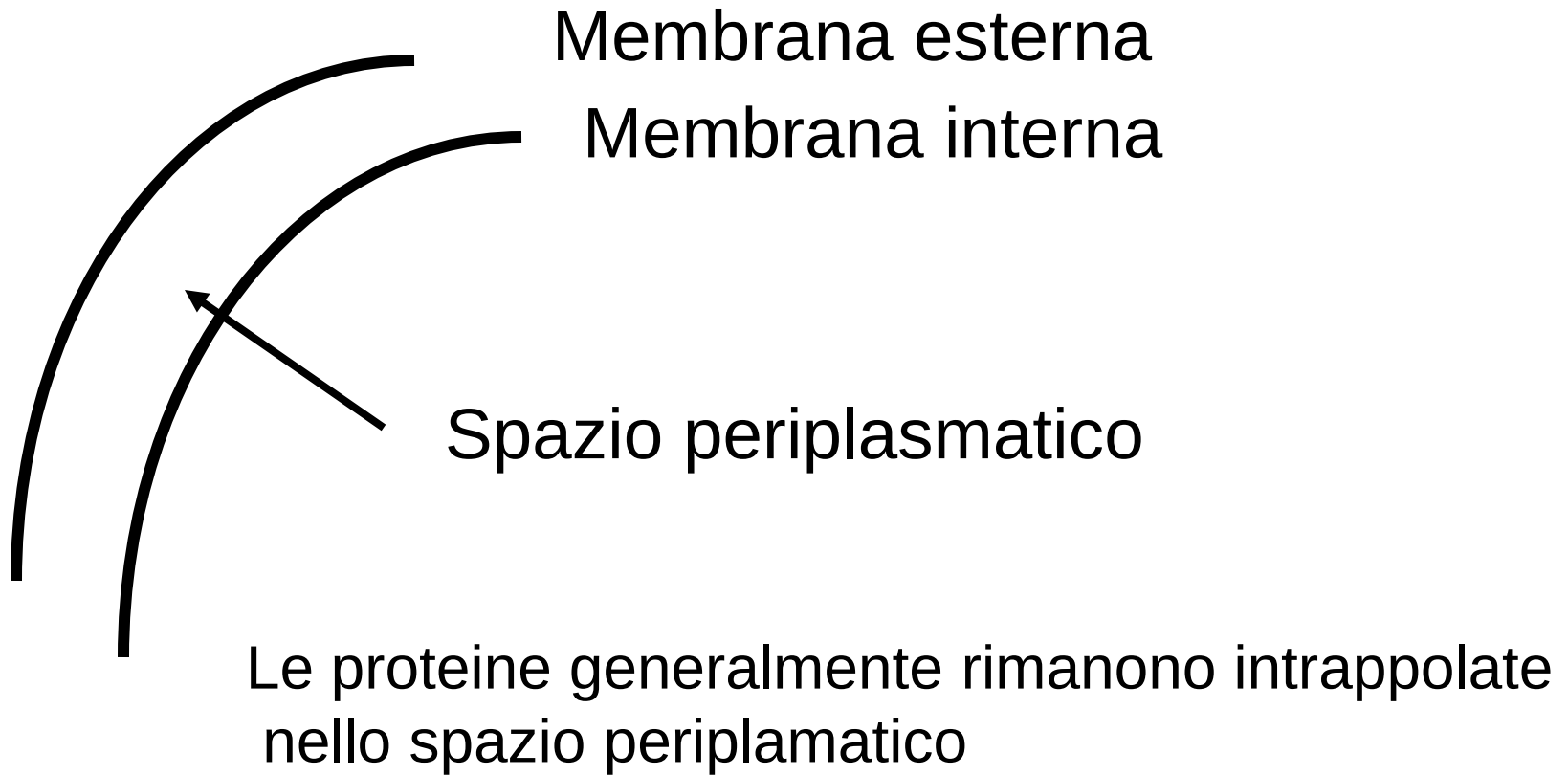
Esporto della Proteina

- Cosa è richiesto per l'esporto?
 - Peptide segnale
 - Questo va aggiunto all'estremità NH_2 della proteina (5' del gene)



↖ Deve essere in frame!

Esporto nei batteri Gram-



Stabilità della proteina

- La vita media è variabile
 - Da pochi minuti a >24 ore
- Fattori che influenzano la stabilità
 - NH₂ terminale
 - Sequenze PEST

Stabilità della proteina

- Sequenze PEST
 - Interne alla sequenza aminoacidica
 - Aumentano la suscettibilità alla proteolisi
 - P = Pro
 - E = Glu
 - S = Ser
 - T = Thr

Table 6.4 Stability of β -galactosidase with certain amino acids added to its N terminus

Amino acid added	Half-life
Met, Ser, Ala	>20 h
Thr, Val, Gly	>20 h
Ile, Glu	>30 min
Tyr, Gln	~10 min
Pro	~7 min
Phe, Leu, Asp, Lys	~3 min
Arg	~2 min

Adapted from Bachmair et al., 1986, *Science* **234**: 179–186.

Proteine di fusione

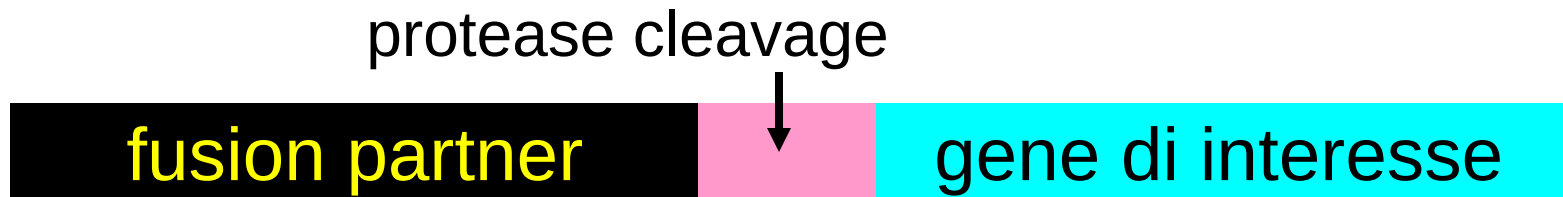
- Ci sono diversi problemi con le proteine, tra cui:
 - Degradazione
 - Purificazione

Proteine di fusione

- La degradazione porta a bassi livelli di espressione e difficoltà nel purificare la proteina.
- Una soluzione frequente:
 - Attaccare la proteina a un partner proteico stabile. Può essere utile rimuovere il “Tag”

Proteine di fusione

- Servono proteasi per separare le due proteine
 - Se ne conoscono alcune che riconoscono una sequenza altamente specifica
 - Tale sequenza va inserita tra la proteina di interesse e il partner proteico



Proteine di fusione

- L'uso primario è per la purificazione
 - In questo caso si usano specifici “tags” per facilitare la purificazione, possibilmente tramite un unico passaggio cromatografico (ad es. mediante cromatografia per affinità)

Proteine con His-tag

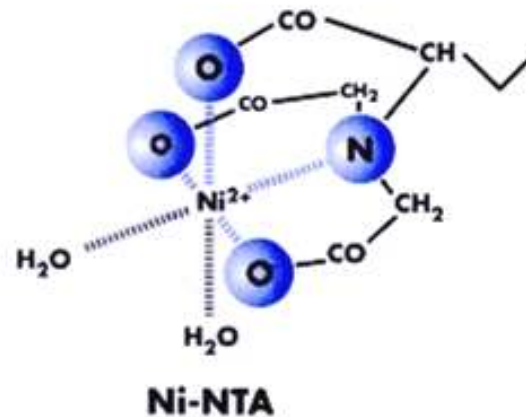
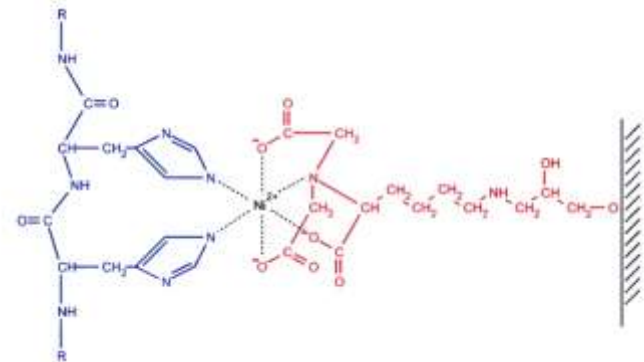
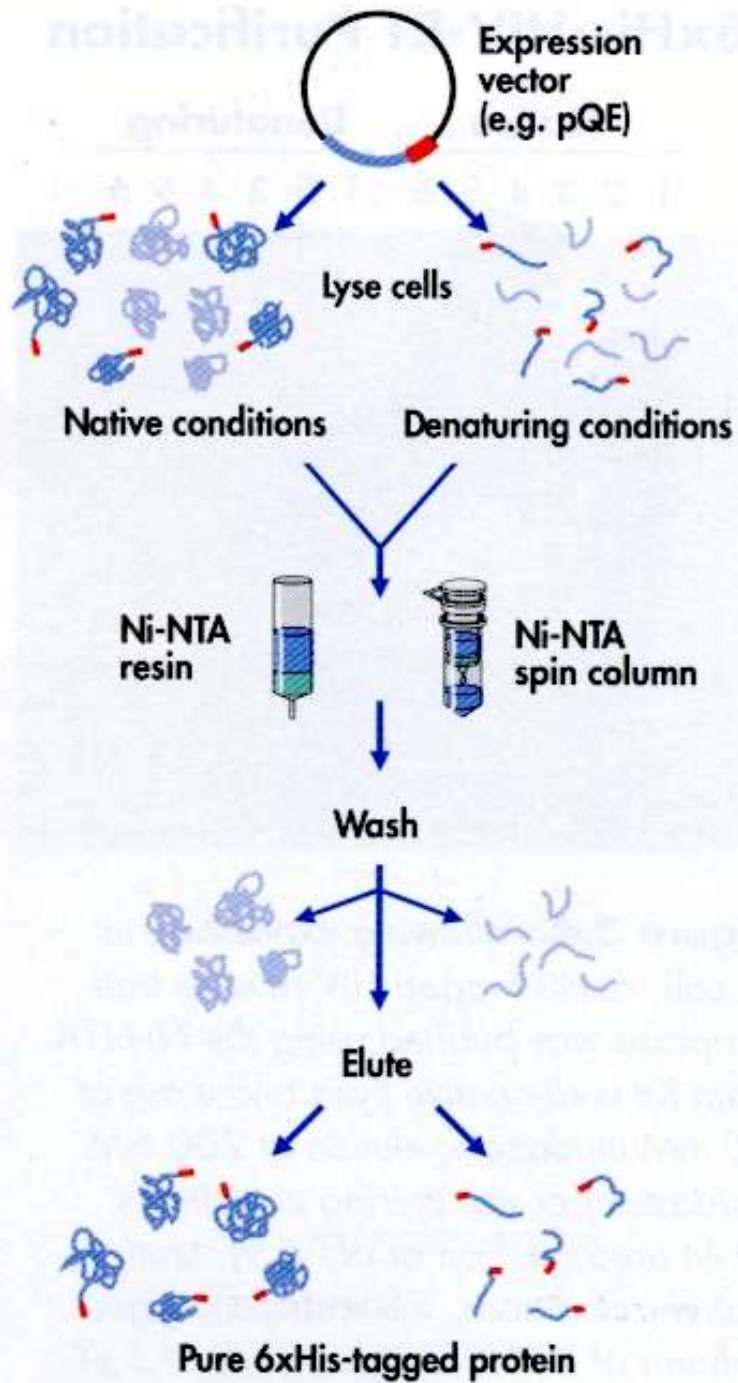


Figure 4. Comparison of the interactions of different metal chelate matrices with nickel ions.



Ulteriori strategie per favorire il “folding” delle proteine

- Co-espressione di Chaperoni
- Modificazione dei ceppi batterici (knockout di geni per proteasi, etc..)
- Modificazioni mirate delle proteine
-

Inclusion Bodies

- Gli inclusion bodies possono essere purificati facilmente
- “Refolding” in vitro, Denaturazione-rinaturazione

Sintesi in sistemi cell-free

E' possibile far sintetizzare proteine in estratti cellulari di origine procariotica (*E. coli*) od eucariotica (germe di grano)

- Nei casi più favorevoli consente di ottenere un'alta efficienza di produzione in un volume molto piccolo, che facilita la successiva purificazione della proteina
- Nessun problema di tossicità
- Consente l'incorporazione di amino acidi marcati per studi metabolici o strutturali
- Possono essere inseriti aminoacidi non naturali
- Più semplice controllare l'azione di proteasi o modulare il folding
- Può essere adattato a produrre proteine con o senza modificazioni post-traduzionali