

# **Produzione di proteine ricombinanti**

# Applicazioni principali delle proteine ricombinanti

- STUDIO DELLA RELAZIONE TRA STRUTTURA E ATTIVITA'
- .-PROTEINE DI INTERESSE TERAPEUTICO (anticorpi, insulina, ormone crescita).
- PROTEINE DI INTERESSE COMMERCIALE (enzimi).
-

Le proteine ricombinanti possono essere espresse come

- *Wild type* equivalenti alle proteine naturali
- Mutanti varianti nelle quali sono state inseriti dei cambiamenti nella sequenza mediante tecniche di mutagenesi

## Espressione di una proteina ricombinante

1. Clonaggio del gene o del cDNA (screening di librerie di cDNA, RT-PCR, RACE, TAIL PCR)
2. Inserimento nel vettore di espressione (promotore inducibile o costitutivo)
3. Trasformazione dell'ospite (shock termico, elettroporazione, liposomi)
4. Selezione dei ricombinanti (resistenza ad antibiotici, vantaggio selettivo)
5. Espressione e purificazione della proteina (intracellulare o secreta)

## Strategia di clonaggio di sequenze codificanti note

- Analisi di restrizione del cDNA/gene che deve essere clonato
- Scelta e del vettore ed analisi del suo MCS
- Progettazione oligonucleotidi contenenti estensioni con i siti di restrizione scelti
- Amplificazione per PCR o RT pCR del DNA da clonare
- Digestione del frammento amplificato e del vettore con gli enzimi di restrizione
- Purificazione dei frammenti
- Ligazione e trasformazione



## Strategia di clonaggio di sequenze codificanti ignote

- Screening di librerie genomiche o di cDNA
- RACE Rapid Amplification of cDNA ends
- TAIL PCR

### Requisito necessario

Conoscenza di alcune porzioni della sequenza amminoacidica per sintetizzare oligonucleotidi da utilizzare nelle reazioni di PCR

## Progettazione di oligonucleotidi a partire da una sequenza amminoacidica

Si sintetizzano delle miscele di oligonucleotidi in modo tale che sia presente almeno un oligo che possa appaiarsi allo stampo



## Il codice genetico é degenerato

		Seconda base del codone					
		T	C	A	G		
Prima base del codone	T	TTT → Phe <b>F</b>	TCT → Ser <b>S</b>	TAT → Tyr <b>Y</b>	TGT → Cys <b>C</b>	Terza base del codone	<b>T</b>
		TTC → Phe <b>F</b>	TCC → Ser <b>S</b>	TAC → Tyr <b>Y</b>	TGC → Cys <b>C</b>		<b>C</b>
		TTA → Leu <b>L</b>	TCA → Ser <b>S</b>	TAA → <b>STOP</b>	TGA → <b>STOP</b>		<b>A</b>
		TTG → Leu <b>L</b>	TCG → Ser <b>S</b>	TAG → <b>STOP</b>	TGG → Trp <b>W</b>		<b>G</b>
	C	CTT → Leu <b>L</b>	CCT → Pro <b>P</b>	CAT → His <b>H</b>	CGT → Arg <b>R</b>		<b>T</b>
		CTC → Leu <b>L</b>	CCC → Pro <b>P</b>	CAC → His <b>H</b>	CGC → Arg <b>R</b>		<b>C</b>
		CTA → Leu <b>L</b>	CCA → Pro <b>P</b>	CAA → Gln <b>Q</b>	CGA → Arg <b>R</b>		<b>A</b>
		CTG → Leu <b>L</b>	CCG → Pro <b>P</b>	CAG → Gln <b>Q</b>	CGG → Arg <b>R</b>		<b>G</b>
	A	ATT → Ile <b>I</b>	ACT → Thr <b>T</b>	AAT → Asn <b>N</b>	AGT → Ser <b>S</b>		<b>T</b>
		ATC → Ile <b>I</b>	ACC → Thr <b>T</b>	AAC → Asn <b>N</b>	AGC → Ser <b>S</b>		<b>C</b>
		ATA → Ile <b>I</b>	ACA → Thr <b>T</b>	AAA → Lys <b>K</b>	AGA → Arg <b>R</b>		<b>A</b>
		ATG → Met <b>M</b>	ACG → Thr <b>T</b>	AAG → Lys <b>K</b>	AGG → Arg <b>R</b>		<b>G</b>
	G	GTT → Val <b>V</b>	GCT → Ala <b>A</b>	GAT → Asp <b>D</b>	GGT → Gly <b>G</b>		<b>T</b>
		GTC → Val <b>V</b>	GCC → Ala <b>A</b>	GAC → Asp <b>D</b>	GGC → Gly <b>G</b>		<b>C</b>
		GTA → Val <b>V</b>	GCA → Ala <b>A</b>	GAA → Glu <b>E</b>	GGA → Gly <b>G</b>		<b>A</b>
		GTG → Val <b>V</b>	GCG → Ala <b>A</b>	GAG → Glu <b>E</b>	GGG → Gly <b>G</b>		<b>G</b>

## COME PROGETTARE UN OLIGONUCLEOTIDE A PARTIRE DA UNA SEQUENZA AMINOACIDICA

- Scegliere la sequenza aminoacidica a minore degenerazione
- Escludere la terza base dell'ultimo codone dal conteggio
- Inserire la deossiinosina
- Considerare il «Codon usage»

Met-Leu-Pro-Gln-Lys-Trp-Asp-Gly-Ser-Met-Asn-Gly-Arg  
1 6 4 2 2 1 2 4 6 1 2 4 6

Pro-Gln-Lys-Trp-Asp-Gly  
4 2 2 1 2 4

CCN-CAA/G-AAA/G-TGG-GAT/C-GGN 18 basi degenerazione 128  
17 basi degenerazione 32

Met-Leu-Pro-Gln-Lys-Trp  
1 6 4 2 2 1

ATG-TTA/G-CCN-CAA/G-AAA/G-TGG 18 basi degenerazione 32 (96)

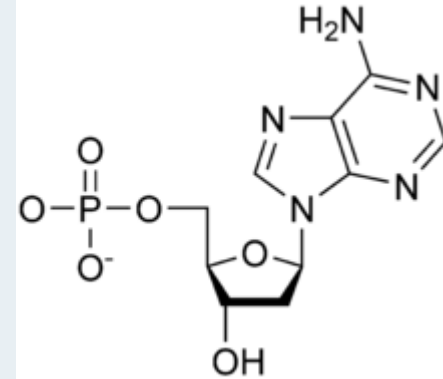
# COME PROGETTARE UN OLIGONUCLEOTIDE A PARTIRE DA UNA SEQUENZA AMINOACIDICA

## DEOSSIIINOSINA

I → A

I → C

I → T



# Frequenza di codoni usati in *Saccharomyces cerevisiae*

*Saccharomyces cerevisiae* [gbpln]: 11937 CDS's (5695112 codons)

fields: [triplet] [amino acid] [fraction] [frequency: per thousand] ([number])

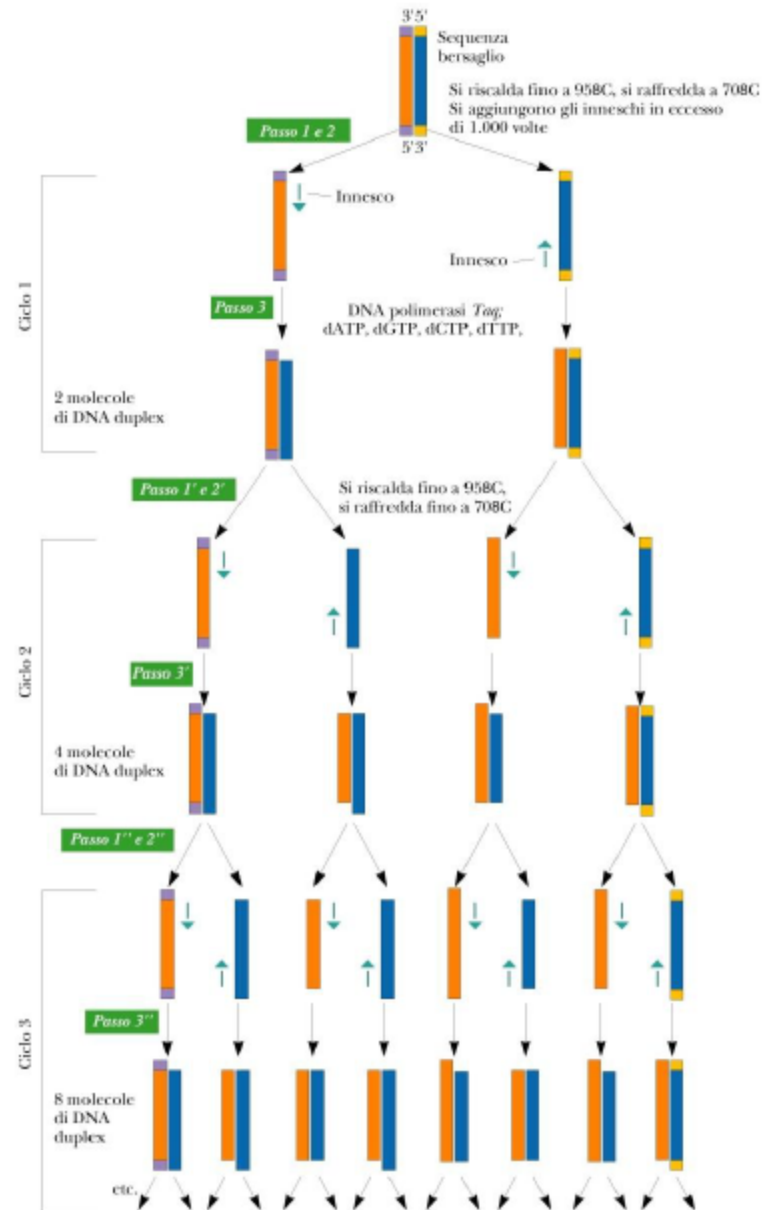
UUU F 0.59 26.1 (148716)	UCU S 0.26 23.5 (134028)	UAU Y 0.56 18.9 (107653)	UGU C 0.63 8.0 ( 45462)
UUC F 0.41 18.2 (103686)	UCC S 0.16 14.2 ( 80799)	UAC Y 0.44 14.6 ( 83397)	UGC C 0.37 4.7 ( 26534)
UUA L 0.28 26.7 (151850)	UCA S 0.21 18.8 (106932)	UAA * 0.47 1.0 ( 5692)	UGA * 0.31 0.7 ( 3804)
UUG L 0.28 27.0 (153619)	UCG S 0.10 8.5 ( 48582)	UAG * 0.22 0.5 ( 2682)	UGG W 1.00 10.3 ( 58581)
CUU L 0.13 12.1 ( 69114)	CCU P 0.31 13.6 ( 77417)	CAU H 0.64 13.7 ( 78282)	CGU R 0.15 6.5 ( 36777)
CUC L 0.06 5.4 ( 30538)	CCC P 0.15 6.8 ( 38522)	CAC H 0.36 7.7 ( 44112)	CGC R 0.06 2.6 ( 14638)
CUA L 0.14 13.3 ( 75972)	CCA P 0.42 18.2 (103680)	CAA Q 0.69 27.5 (156586)	CGA R 0.07 3.0 ( 17099)
CUG L 0.11 10.4 ( 59114)	CCG P 0.12 5.3 ( 29915)	CAG Q 0.31 12.1 ( 69006)	CGG R 0.04 1.7 ( 9848)
AUU I 0.47 30.4 (173137)	ACU T 0.35 20.3 (115551)	AAU N 0.59 36.3 (206516)	AGU S 0.16 14.2 ( 80728)
AUC I 0.26 17.0 ( 97055)	ACC T 0.21 12.6 ( 71608)	AAC N 0.41 24.9 (141737)	AGC S 0.11 9.7 ( 54958)
AUA I 0.27 17.8 (101628)	ACA T 0.30 17.8 (101336)	AAA K 0.58 42.2 (240515)	AGA R 0.48 21.3 (121292)
AUG M 1.00 20.9 (119008)	ACG T 0.14 7.9 ( 45120)	AAG K 0.42 30.7 (174597)	AGG R 0.21 9.2 ( 52441)
GUU V 0.39 22.0 (125255)	GCU A 0.38 21.1 (120208)	GAU D 0.65 37.8 (215249)	GGU G 0.47 24.0 (136434)
GUC V 0.21 11.6 ( 65876)	GCC A 0.22 12.5 ( 71233)	GAC D 0.35 20.2 (115047)	GGC G 0.19 9.7 ( 55103)
GUA V 0.21 11.8 ( 67408)	GCA A 0.29 16.1 ( 91967)	GAA E 0.71 45.8 (261115)	GGA G 0.22 10.9 ( 62176)
GUG V 0.19 10.6 ( 60537)	GCG A 0.11 6.1 ( 34942)	GAG E 0.29 19.1 (108666)	GGG G 0.12 6.0 ( 34032)

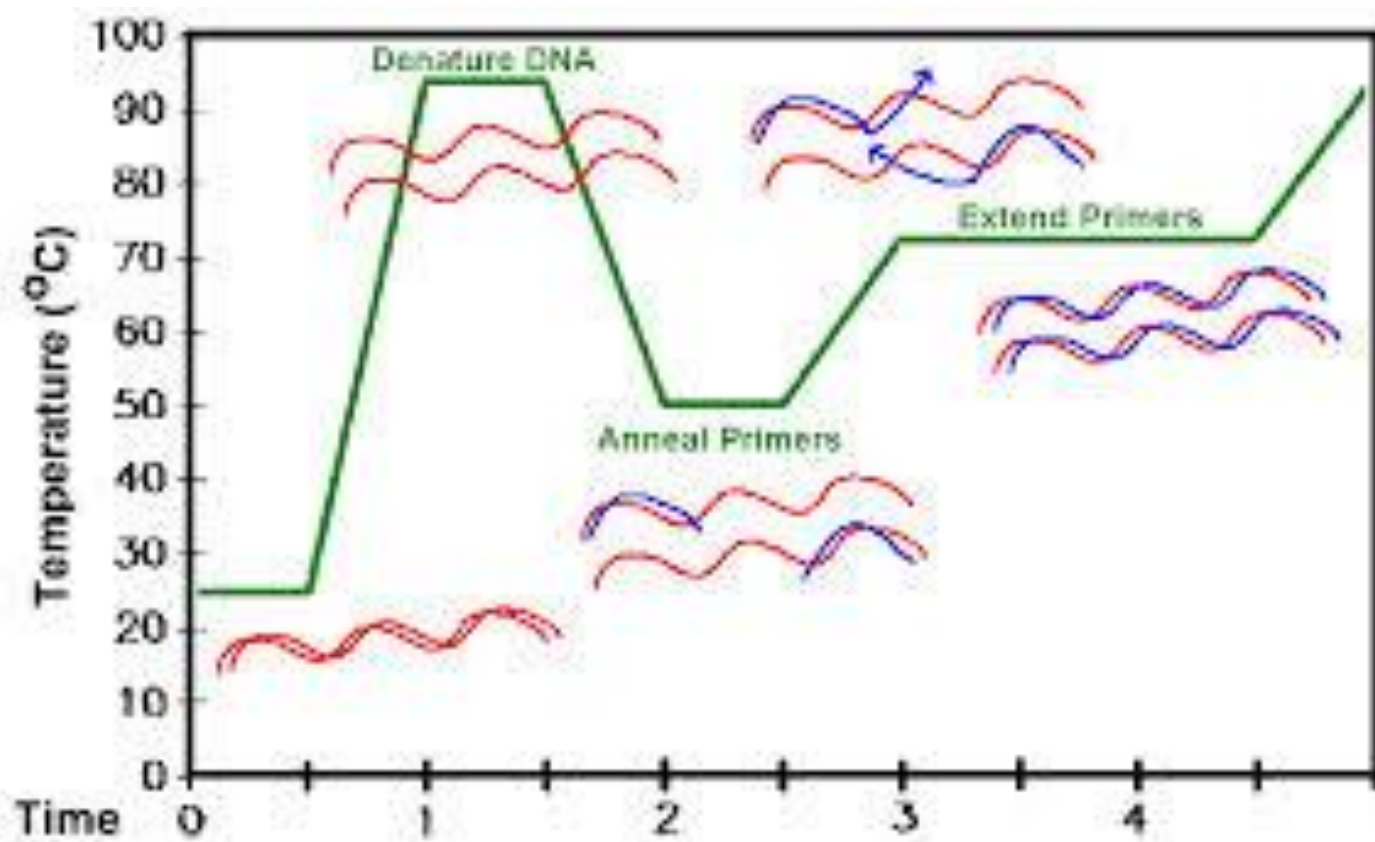
Coding GC 39.64% 1st letter GC 44.53% 2nd letter GC 36.55% 3rd letter GC 37.85%

Genetic code 1: Standard

# PCR (Polymerase Chain Reaction)

**Figura 9.20** La reazione a catena della polimerasi (PCR). Gli oligonucleotidi complementari ad una data sequenza di DNA ne innescano la sintesi. La DNA polimerasi Taq termostabile sopravvive a diversi cicli di riscaldamento. Teoricamente la quantità di sequenza specificata dai primer è raddoppiata ad ogni ciclo.





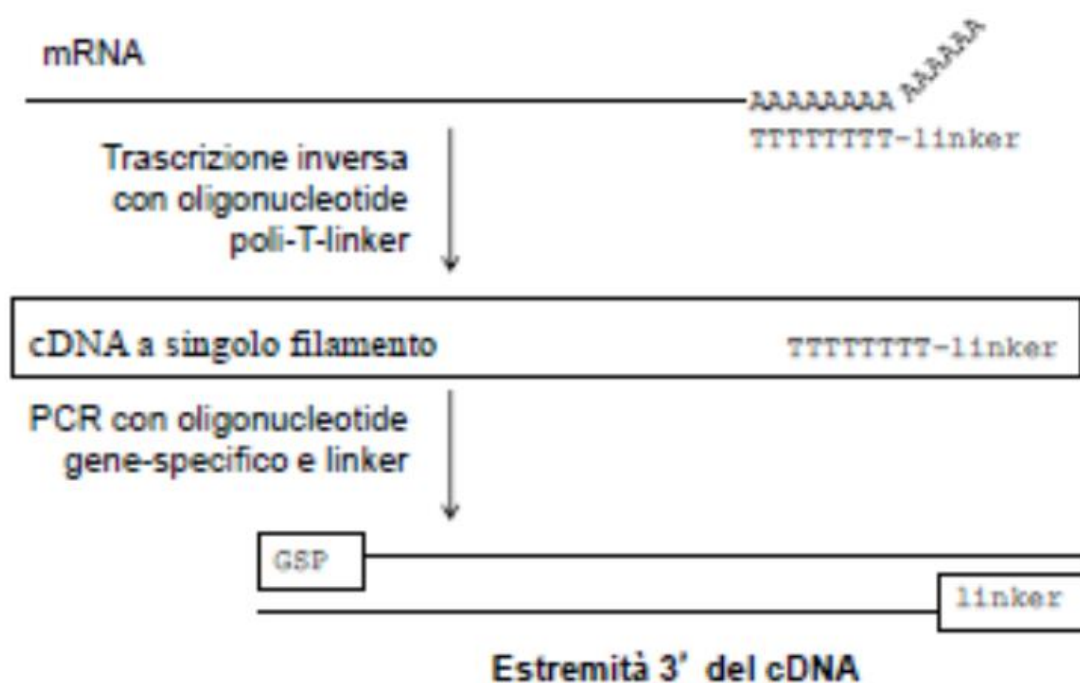
# APPLCAZIONI DELLA PCR

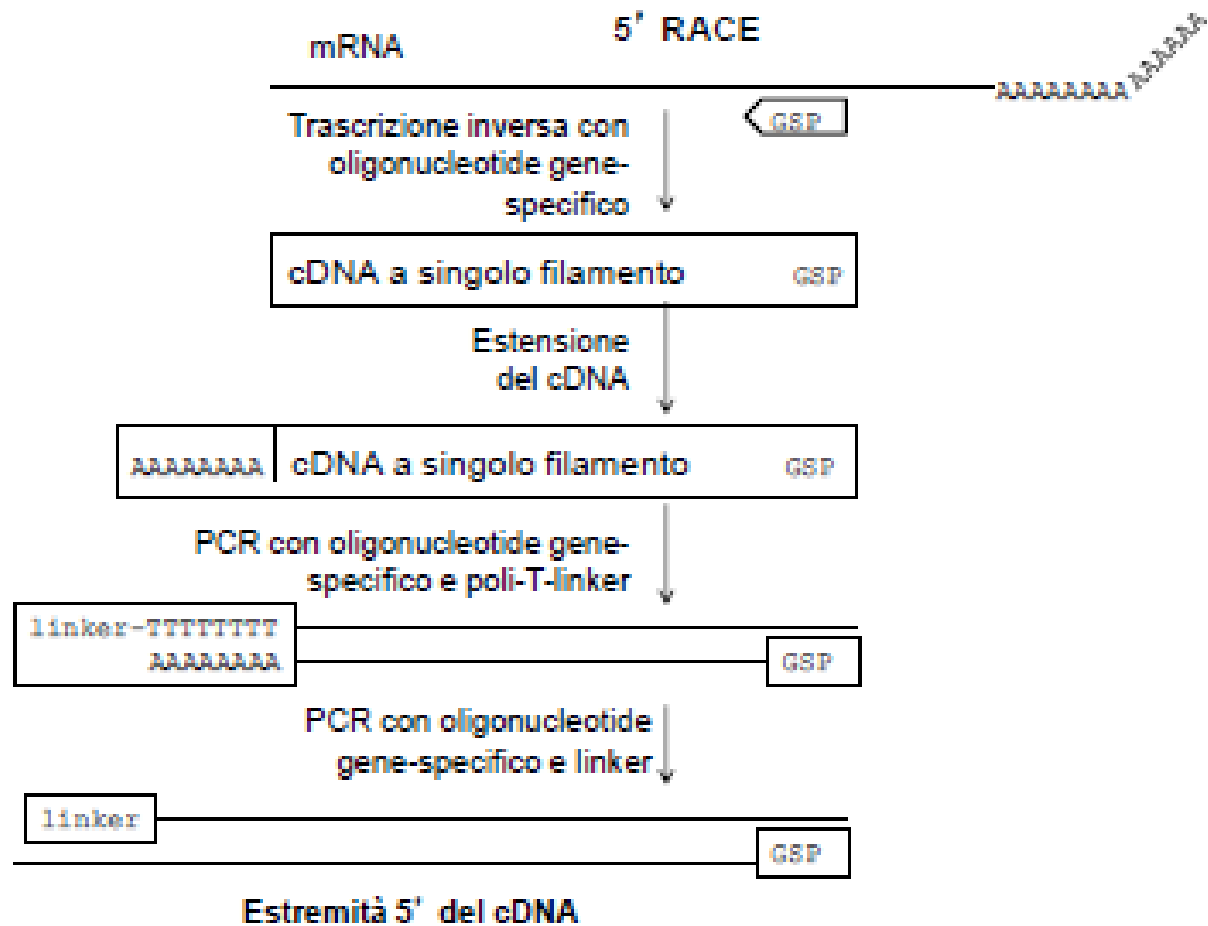
## RACE Rapid Amplification of cDNA ends

l'mRNA viene trascritto in cDNA dalla trascrittasi inversa utilizzando un innesco generico (poli-T+adattatore, 3' RACE) o gene-specifico (5' RACE). Il cDNA viene poi amplificato con l'oligo adattatore e un oligo gene-specifico.



### 3' RACE



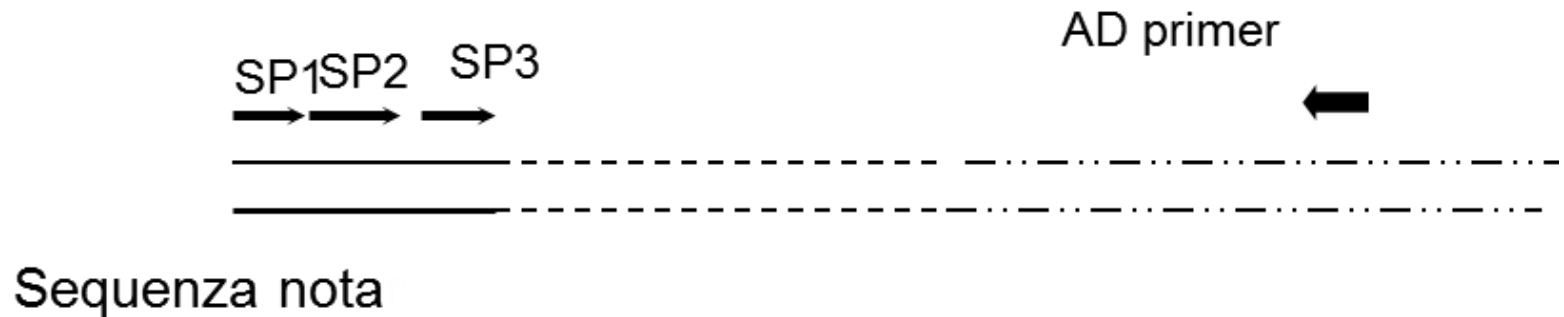


## TAIL PCR (THERMAL ASYMMETRIC INTERLACED PCR)

TAIL-PCR è una tecnica che permette di amplificare frammenti di DNA adiacenti a sequenze note

- consiste in tre reazioni consecutive di PCR in cui vengono utilizzati tre primer specifici consecutivi SP1 SP2 e SP3 e un primer arbitrario AD degenerato avente una  $T_m$  più bassa

le amplificazioni specifiche e aspecifiche sono controllate termicamente



# Applicazioni

- Isolamento regioni codificanti presenti al 5'
- Isolamento di promotori
- Amplificazione di sequenze poco rappresentate da genomi complessi

## TAIL-PCR

### Progettazione primers

#### **Oligonucleotidi specifici (SP1-SP2-SP3):**

- *Nested*
- Complementari alla sequenza codificante
- $T_m = 58-63^\circ\text{C}$

#### **Oligonucleotidi degenerati :**

- Corti
- $T_m$  più bassa di quella degli oligo specifici ( $T_m = 47-48^\circ\text{C}$ )

### Temperatura di annealing

variabile

### Ordine dei cicli

si alternano cicli ad alta e bassa stringenza

-

-

## Temperatura di *annealing*

Cicli ad alta stringenza (thermal asymmetric)

Annealing temperature = 63°C

Cicli a ridotta stringenza (thermal symmetric)

Annealing temperature = 44°C

Cicli a bassa stringenza

Annealing temperature = 30°C

**Cycling Conditions Used for TAIL-PCR  
on the GeneAmp System 9600**

Reaction	File no.	Cycle no.	Thermal condition
Primary	1	1	92°C (2 min), 95°C (1 min)
	2	5	94°C (15 s), 63°C (1 min), 72°C (2 min)
	3	1	94°C (15 s), 30 °C (3 min), ramping to 72°C over 3 min, 72°C (2 min)
	4	10	94°C (5 s), 44°C (1 min), 72°C (2 min)
	5	12 <sup>a</sup>	94°C (5 s), 63°C (1 min), 72°C (2 min) 94°C (5 s), 63°C (1 min), 72°C (2 min) 94°C (5 s), 44°C (1 min), 72°C (2 min)
	6	1	72°C (5 min)
Secondary	7	10 <sup>a</sup>	94°C (5 s), 63°C (1 min), 72°C (2 min) 94°C (5 s), 63°C (1 min), 72°C (2 min) 94°C (5 s), 44°C (1 min), 72°C (2 min)
	6	1	72°C (5 min)
Tertiary	8	20	94°C (10 s), 44°C (1 min), 72°C (2 min)
	6	1	72°C (5 min)

*Note.* The program files in each reaction were linked automatically.

<sup>a</sup> These are nine-segment super cycles each consisting of two high-stringency and one reduced-stringency cycle (see Fig. 1).

I PCR primers  
SP1 e AD

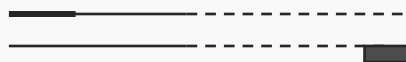
- 5 cicli ad alta stringenza T annealing 63
- 1 ciclo a bassa stringenza T annealing 44
- 10 cicli a ridotta stringenza Tm annealing variabile (ramping)

12 TAIL-cycling  
(super cicli)

1 ciclo a bassa stringenza  
(thermal symmetric)

2 cicli ad alta stringenza  
(thermal asymmetric)

prodotto specifico  
(type I)



Product yield  
High or middle  
(detectable or undetectable)

prodotto non specifico  
(type II)



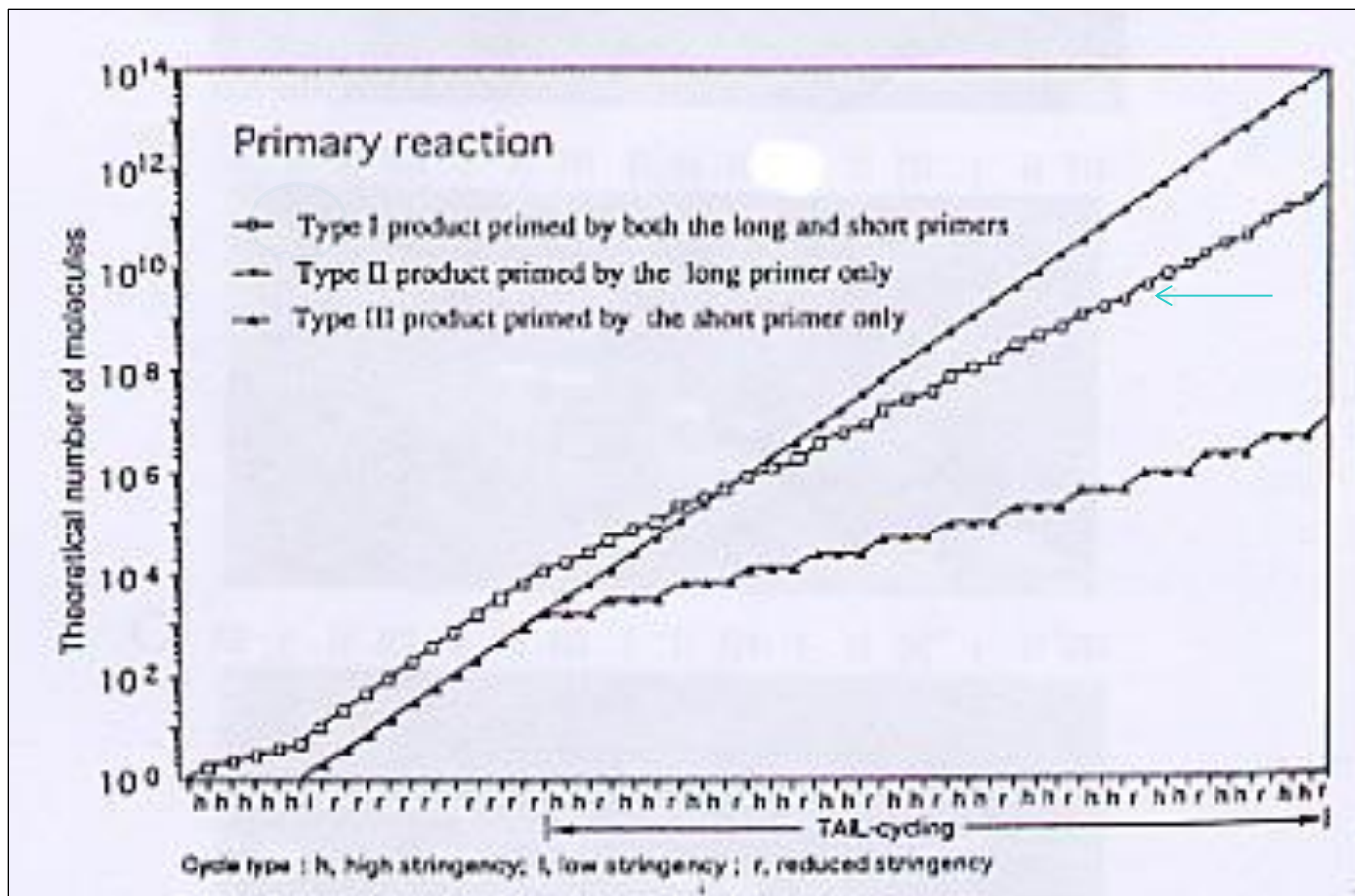
High  
(detectable)

Prodotto non specifico  
(type III)



Low  
(undetectable)





## II PCR

PCR con SP2 e AD (10 super cicli)

templato è il prodotto della prima reazione di PCR diluito 1000-volte



Prodotto specifico



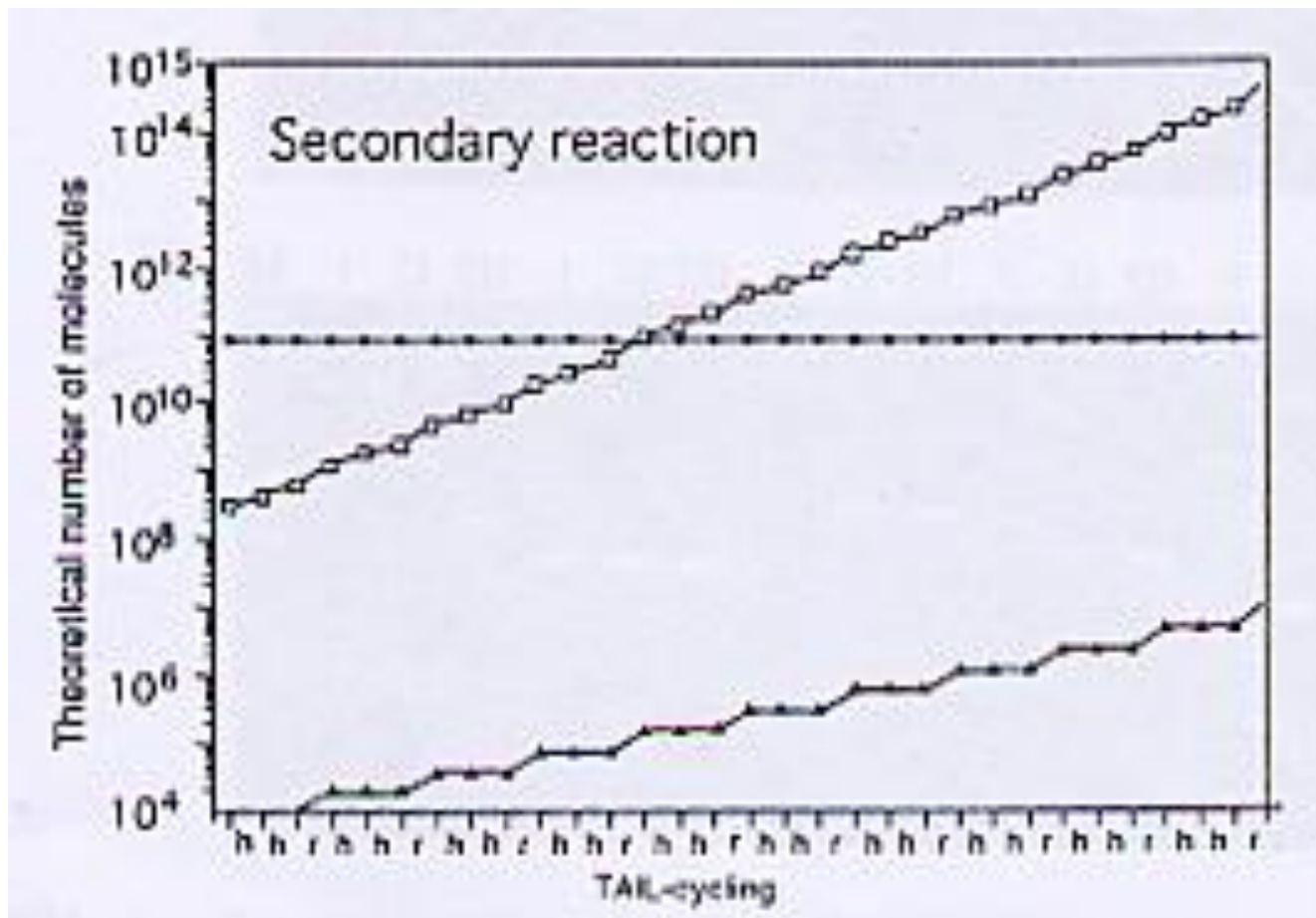
Prodotto nonspecifico (type III)



Product  
yield:

High (detectable)

Very low (undetectable)



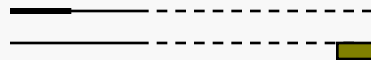
## III PCR

### PCR con SP3 E AD (20 cycles)

- templatò è il prodotto della seconda reazione di PCR diluito 1000-volte



Prodotti specifici



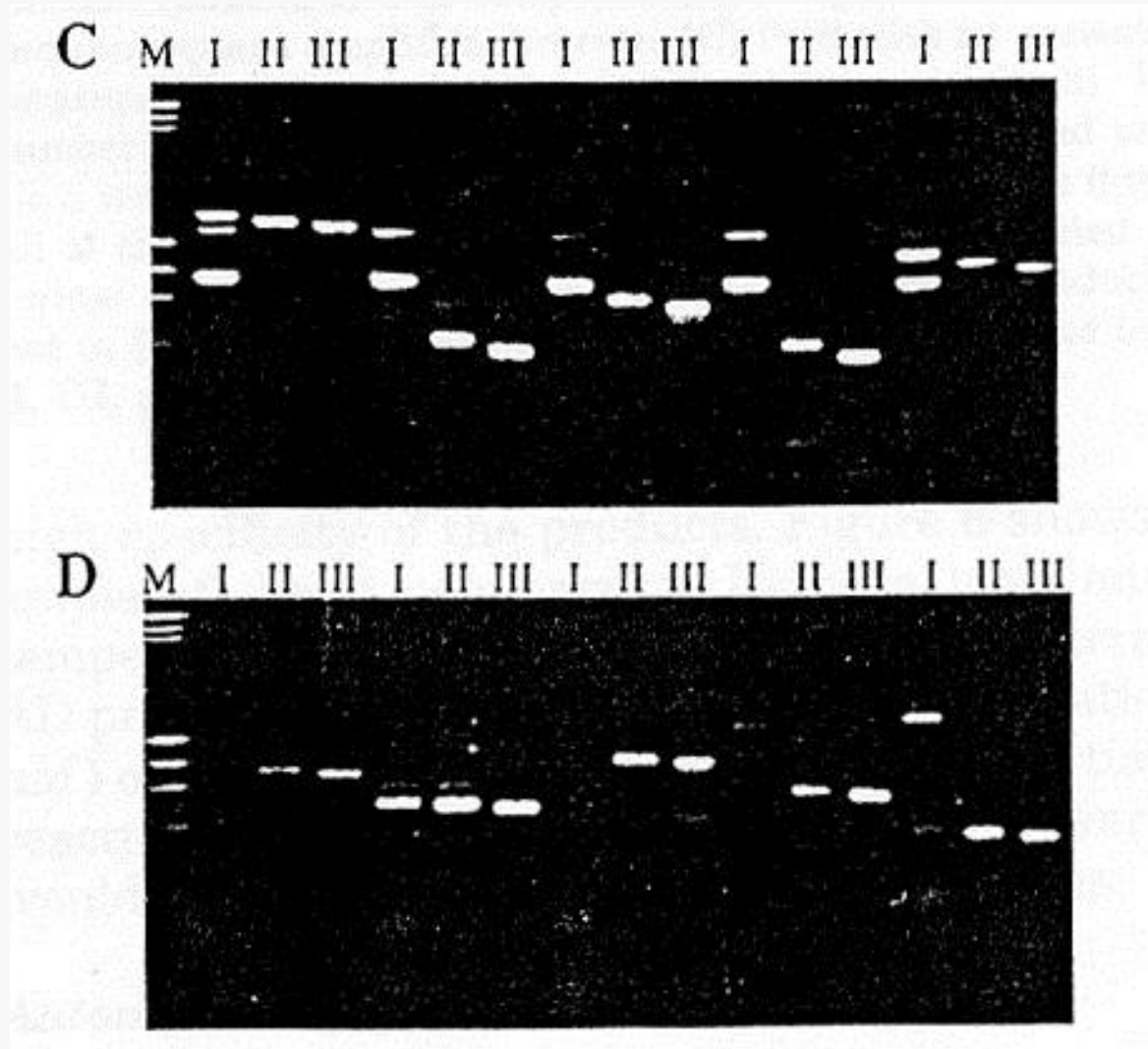
Gel di Agarosio

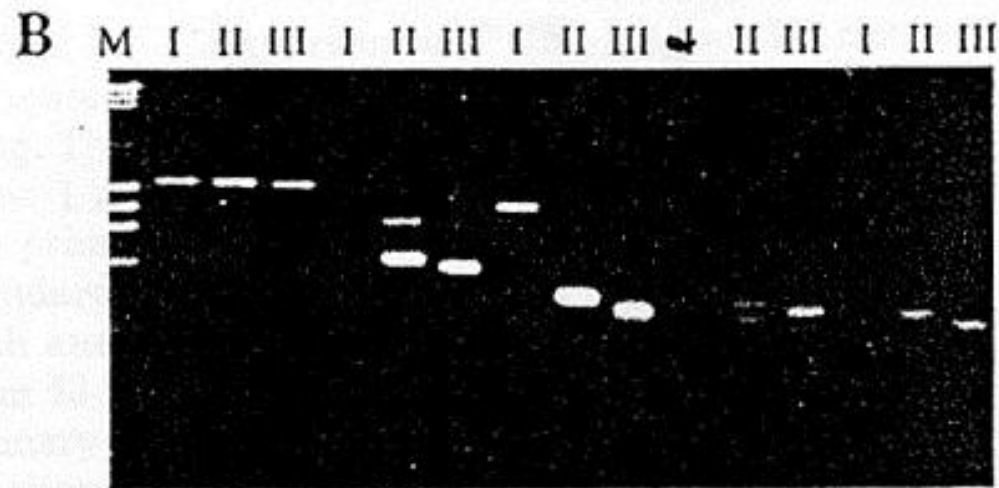
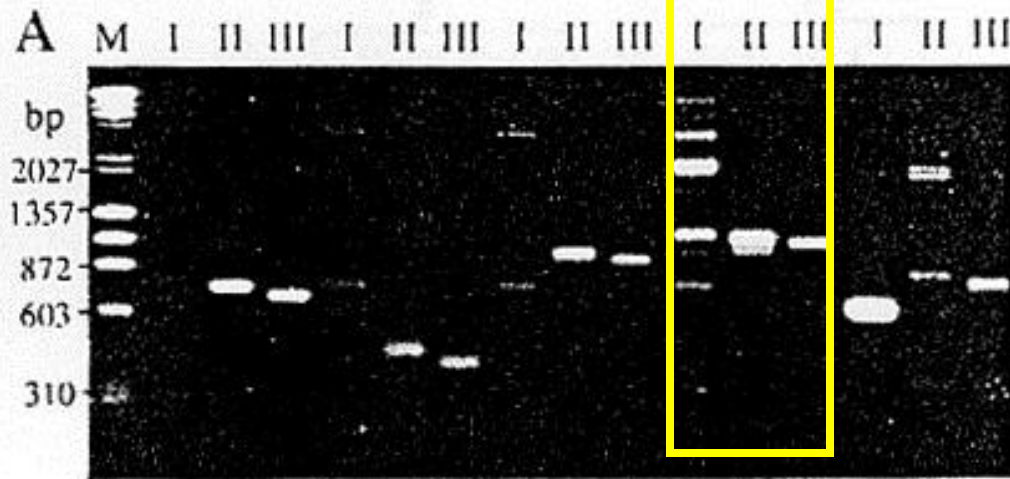


Sequenziamento

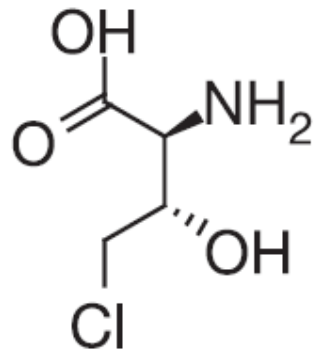
- Molti prodotti di PCR dalla prima TAIL-PCR che scompaiono nella seconda TAIL-PCR
- Prodotti specifici non sono sempre presenti nella prima reazione.

# YAC clones





# Identificazione del cluster biosintetico della 4-clorotreonina in *Streptomyces* sp. OH- 5093



4-Cl-L-Thr

Attività antifungina

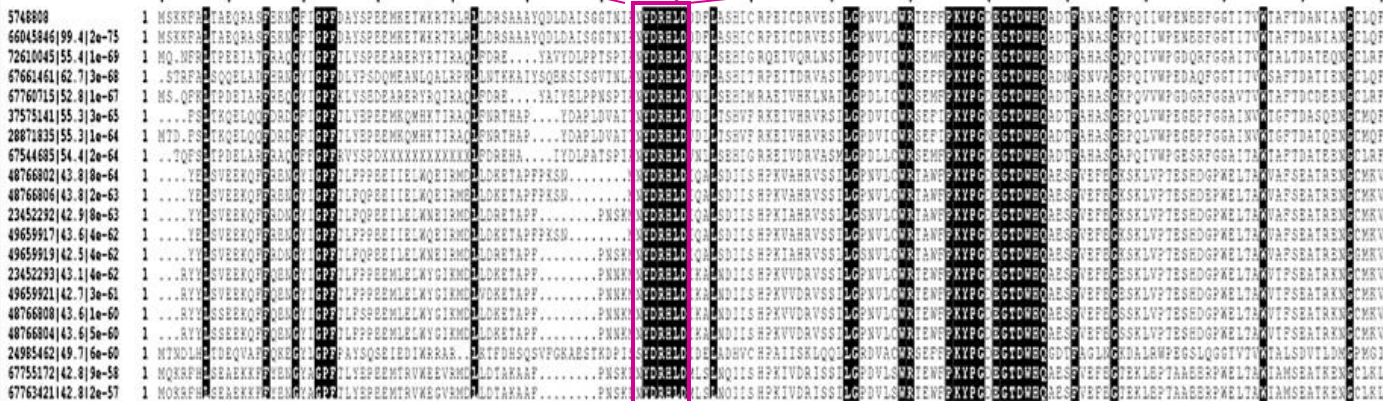




# Identificazione dell'alogenasi in *Streptomyces* sp. OH-5093

## 1. Allineamento proteine ortologhe a SyrB2

$N_{65}-D_{71}$ : 5'- AAC TAC GAC CG(GC) CAC CT(GC) GA(CT)-3'



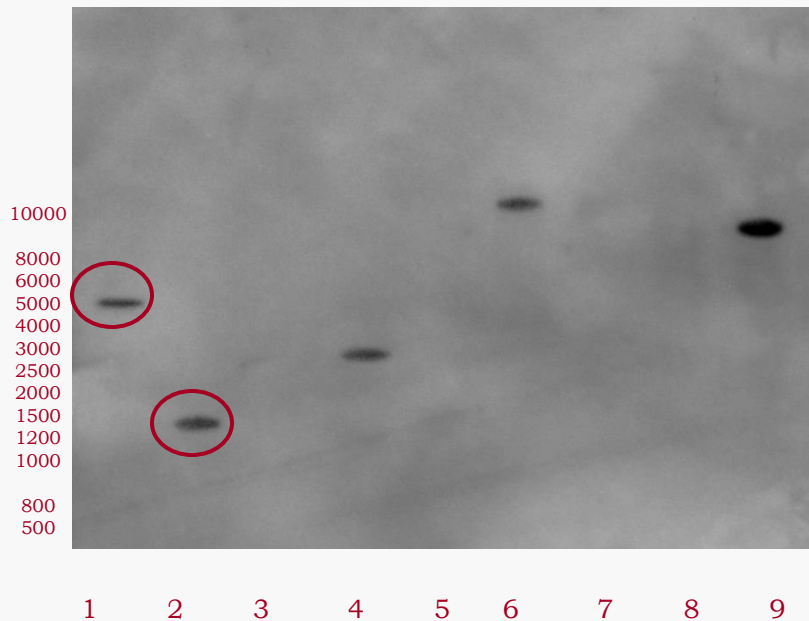
## 2. Scelta delle regioni altamente conservate per la costruzione di oligonucleotidi degenerati

$G_{194}-Y_{200}$ : 3'-CC(CG) (CA)A(CG) A(TA)G CC(CG) ATG CTG ATG-5'

### 3. Amplificazione per PCR e sequenziamento

```
1   N Y D R H L D V D L L S E H I M R E E I  
1   AACTACGACCGCCACCTCGACGTCGACCTGCTCTCGAGCACATCATGAGGGAGGAGATC  
  
21  V D R V V S I L G P D V L C W R T E F F  
61  GTCGACCGGTCGTCTCCATCCTGGGGCCCGACGTCCTGTGCTGGCGCACGGAGTTC TTC  
  
41  P K Y P G D E G T D W H Q A A T F A N A  
121 CCCAAGTACCCGGGTGACGAGGGGACC GACTGGCACCAGGCGGCGACGTTCCGCAACGCC  
  
61  S K K P Q I M W P S E E G R P A F G G T  
181 AGCAAGAAGCCGAGATCATGTGGCCGTCCGAGGAGGGCCGGCCCGCTTTCGGCGGCACC  
  
81  I T V W T A F T D S T L E N G C L Q L M  
241 ATCACC GTGGACCGCTTTCACCGACTCGACGCTGGAGAACGGCTGCCTCCAGCTGATG  
  
101 P G T H K H M N Y D E S L A M D Y D P N  
301 CCCGGCACCACAAGCACATGAACTACGACGAGTCGCTGGCCATGGACTACGACCCCAAC  
  
121 A I N S R E K D G V K R G F F G Y D Y  
361 GCCATCAACAGCCGTGAGAAGGACGGCGTCAAGCGGGCTTCTTCGGCTACGACTAC
```

# Analisi di *Southern blotting*

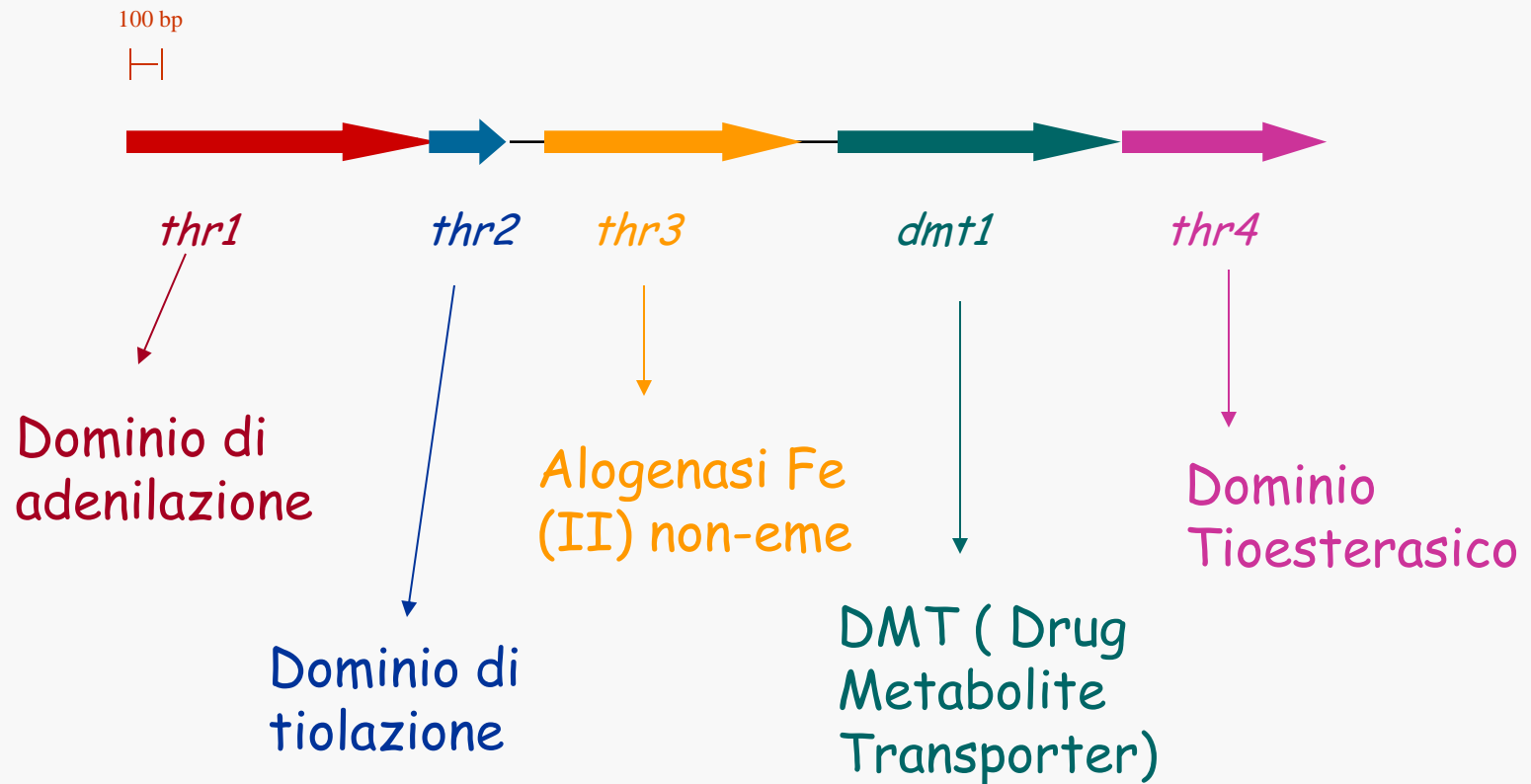


Southern blotting effettuato sul DNA genomico di *Streptomyces* sp. OH-5093 digerito con BamHI (1), PstI (2), RsaI (3), SacII (4), SmaI (5), EcoRI (6), XhoI (7), SphI (8), BfaI (9). La sonda utilizzata è JV2 marcata con  $^{32}\text{P}$ .

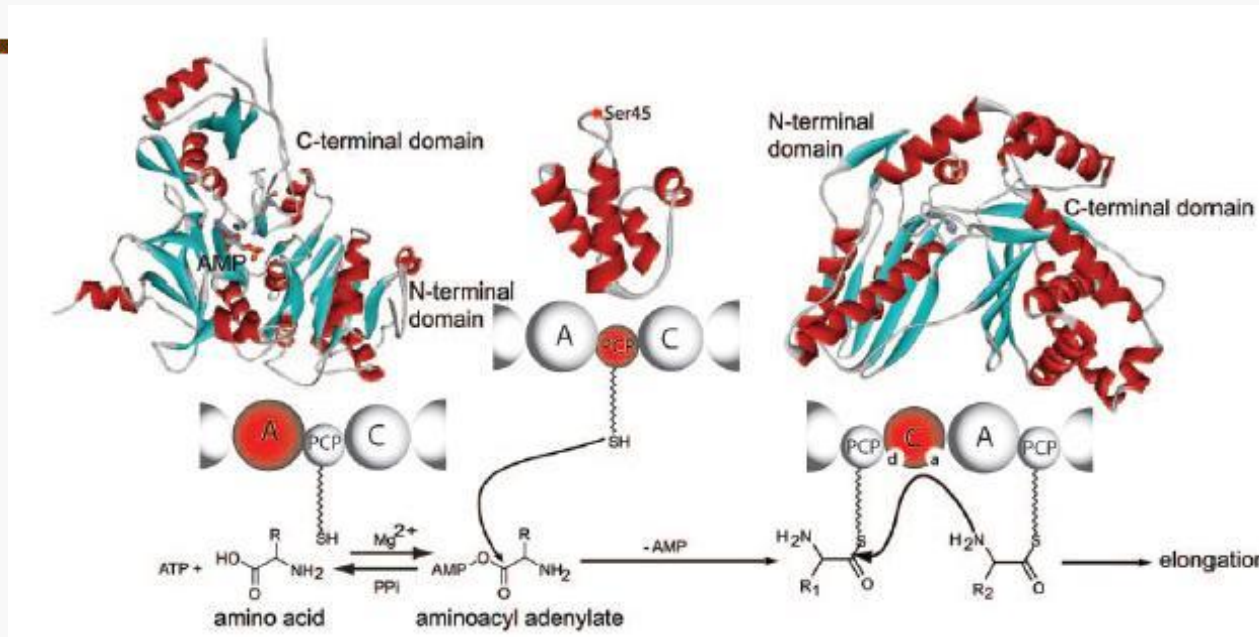
➤ L'alogenasi è presente in singola copia

- Clonaggio dei frammenti in pBlueScript
- Screening su colonia con sonda marcata radioattivamente

# Organizzazione del cluster *thr*



# MECCANISMO DELLA SINTESI PEPTIDICA NON RIBOSOMIALE



- **A** seleziona e attiva l'amminoacido;
- **PCP** lega l'amminoaciladenilato alla 4'-fosfopanteteina, gruppo prostetico del dominio stesso;
- **C** catalizza la formazione del legame peptidico.



**A3**

1 A Y I I Y T S G T T G R P K G V P V R H  
 1 GCCTACATCATCTACACCTCCGGCACACGGCCGCCCAAGGGTCCCGCTCCGGCAC

21 G N V T A L F E A C S R L F S F S A D D  
 61 GGGAAATGTACCCGCTCTCTTCGAGGCCGTGCTCCCGCCTGTTCTCCTTCCTCCGGCCGACGAC

41 R W L L F H S M A F D F S V W E I W G A  
 121 CGCTGGCTGCTCTTCCACTCCATGGCCTTGGACTTCTCGGTCTGGGAGATCTGGGGTGGC

61 L S T G A E L V V L P Y W T A R T P V E  
 181 CTGAGCACCGGCGCGGAACTGGTCGTCTGCGCTACTGGACAGCCCGCACCCCGCTCGAG

81 T A R V V R D R G I T V L N Q T P T A F  
 241 ACCGCGCGAGTCTACGCGACCGGGGATCAGCGTGTGAACACAGACGCGCACCGCCTC

101 G A L T T A V L G E G I D L P E L R Y V  
 301 GGCGCGCTGACGACAGCGTCTCGGGCGAGGGCATCGACCTGCGGAACCTGCGCTACGTC

121 V F G G E K L T P A V V R P W A K R F G  
 361 GTCTTCGGCGGCAGAAAGCTCACGCCCGCGTGGTACGCCCTGGGCCAAGCGTTCGGT

141 L D R P H L I N M Y G I T E T T V H A T  
 421 CTGGACCGGCGCGACCTGATCAACATGTACGGCATCACCGAGACCAGGTGCACGCCACG

161 F H R L T E D D L A A E D S V I G R P L  
 481 TTCACCGGCTGACCGAGGACGACCTGGCAGCGGAGGACTCGGTCATCGGCGGCGCGCTG

181 P G F T H R I V T E D G R D A A T G E P  
 541 CCGCGCTTACCCACCGGATCGTCACCGAGGACGGGCGGACCGCGGACCGGGCGGAGCGC

**A6**

201 G E L W L A G P Q V S E G Y L N R P E L  
 601 GGTGAACCTGTGGCTGGCCGGTCCACAGGTGAGCGAGGGTATCTCAACCGGCCGGAACCT

221 T A E R F T T G P A R D G G A P P P R Y  
 661 ACCGCGAAGCGCTTACCCACCGGTCCCGCGGGACGGCGGGGCGCGCCACCGCGCTAC

**A7**

241 Y H S G D L V S R R A G G D L V Y Q G R  
 721 TACCACTCGGCGACCTGCTCAGCCGGAGCGGGGGCGACCTGGTCTACCAAGGCCGA

**A8**

261 A D L Q V K L R G H R I E L S D V E A A  
 781 GCCGACCTGCAGGTGAAGCTCCGAGGCCATCGCATCGAACTCAGCGATGTGAGGCCGCG

281 V R T H P A V V D A V V W V H E F A P G  
 841 GTGGCGACGCACCCCGGTTGGTCGACGCGGTTGGTGGTGCACGAGTTCGACCGGGC

301 D S R L V C A Y T A Q A S E E G T G P D  
 901 GACAGCCGCTGTGTGGCCTACACAGCACAGGCCCTCCGAGGAGGGCACGGGACCCGAC

321 A R A L R A H V K T V L P S Y M Q P S Q  
 961 GCCGTGCGCTGGCGCCCATGTGAAGACAGTCTGCGCGTGTACATGCAAGCCCTCGCAG

**A9**

341 Y L A L P E L P R T I N G K A D R A S V  
 1021 TATCTGGCGTGGCCGAGCTGCGCGCACGATCAACGGAAAGCCGACCGGGCCTCCGTG

**A10**

361 A R A F D E R R -  
 1081 GCCCGAGCCTTCGACGAAAGGAGATGA

Thr1

Dominio di adenilazione

Presenza dei motivi conservati

Specificità per il riconoscimento della treonina

Identificazione regione 5' del gene thr1  
mediante TAIL PCR




# Thr1

Dominio di adenilazione

 Regioni del promotore

 Possibili codoni d'inizio

 Shine - Dalgarno

 Motivi conservati nei domini di adenilazione

 Primer utilizzati nella TAIL-PCR

```
1  R L G R M R R Q R R P A A I - A A W P L
3  CGACTGGGACGCATGAGGCGACAGCGGCGACCGGCCATTGAGCGGCCTGGCCCCTT

21  L L S G P F L V N S Q R I H R L R F A V
63  TTGCTAAGTGGGCCGTTTCTCGTCAACTCCCAGCGGATTACAGACTCCGATTGCTGTG

41  K T - V Q P E C P N P - R S H V T K I T
123 AAGACATAGGTTTCAGCCAGAGTGTCCAAACCCTTGACGATCACATGTAACTAAGATCACA

61  F I Y G G V R T D I D R V T M - C D F G
183 TTTATCTACGGCGGGTCCGGACAGACATCGATCGTGTAACGATGTAGTGCCTGACTTCGGG

81  P R W T G G F G H L H N G A I V S E K K
243 CCACGTTGGACTGGTGGATTCCGGTCATCTACACAATGGAGCCATTGTGTCCGAAAAGAAG

101  D Q Y P S P G Y S E F G A T L V D L F S
303 GACCAGTACCCCTCGCCAGGATATTCGGAATTCGGCGCGACACTGGTTGACTTGTTCCTCC

121  R A A M E M P D R T A L H I D D E K I S
363 AGGGCAGCCATGGAAATGCCCGACCGCACGGCGCTCCACATCGACGACGAGAAGATCTCG
      A1
141  Y G L I H S W A E G L A D L L H D A G V
423 TACGGCCTCCTCCATTCTGGGCGGAGGGCCTCGCGGATCTTCTCCACGACGCCGGGGTA

161  R K G D R V A L R M P P G A N A I A A M
483 CGGAAGGGTGATCGTGTGCGACTGCGTATGCCGCGGGGCGGAACGCGATCGCCGCCATG
      A2
181  L G I L R A G A A Y V P L D I R N P P A
543 CTGGGAATCCTCCGGGCCGGCGCCGCTATGTTCCGCTCGACATTGCAACCCACCCGCT

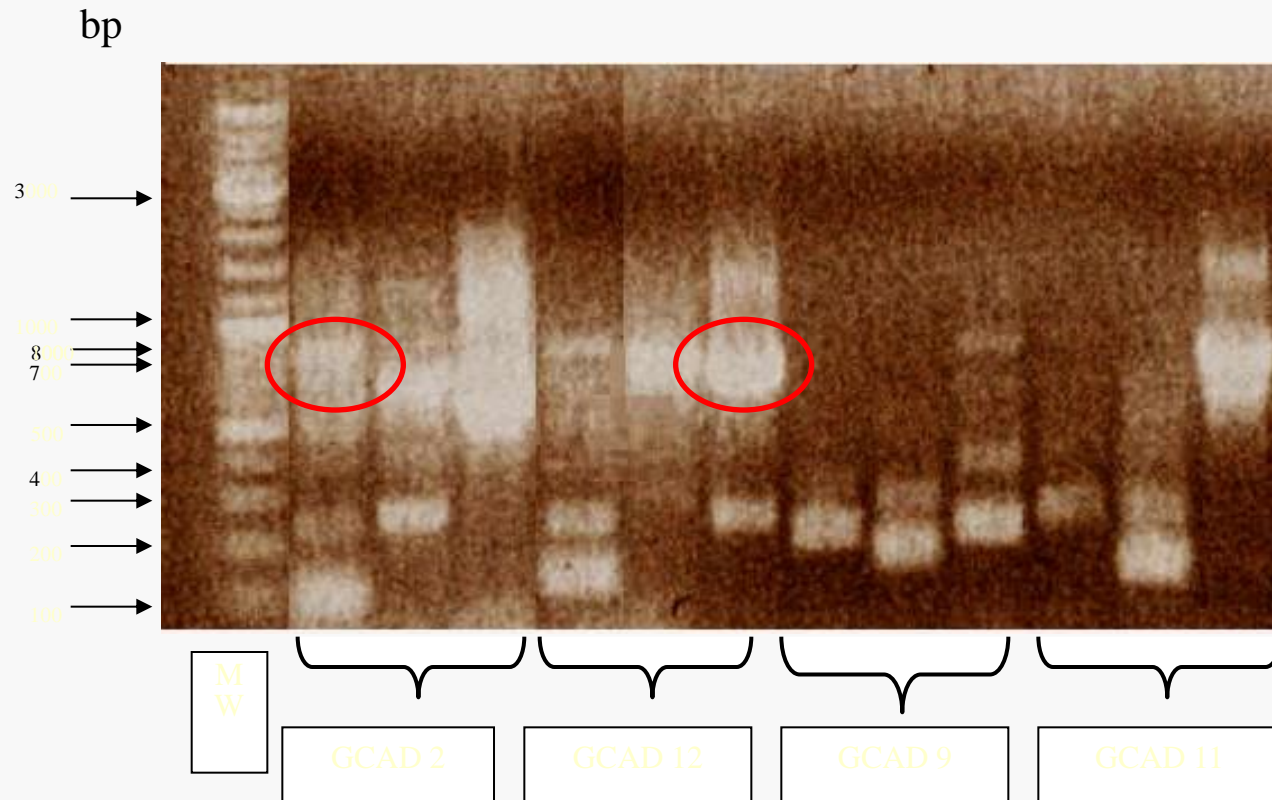
201  R N A F I V T D S Q V V A L V G D P G D
603 CGCAACGCCTTCATCGTCCCGACAGCCAGGTGGTCCGCGTGGTGGGAGATCCCGGTGAT

221  I P E Y T G P L V T E E N V A A L R D R
663 ATTCCCGAATACAGGGGCCGCTCGTCCAGGAGAACGTGGCAGCACTGCGGGACCGT

241  E A I K D P G P E R P G P Q D V A Y I I
723 GAGGCGATCAAGGATCCCGGGCCCGAGCGGCCCGCAGGACGTGCGCTACATCATC
      A3
261  Y T S G T T G R P K G V P V R H G N V T
783 TACACCTCCGGCACACGGGCCGCCCAAAGGGTCCCCGTCCGGCACGGGAATGTCACC
      sp3
281  A L F E A C S R L F S F S A D D R W L L
843 GCTCTCTTCGAGGCCTGCTCCCGCTGTTCTCCTTCTCCGCGGACGACCGCTGGCTGCTC
      A4 sp2
301  F H S M A F D F S V W E I W G A L S T G
903 TTCCACTCCATGGCCTTCGACTTCTCGGTCTGGGAGATCTGGGGTGCCTGAGCACCGGC

321  A E L V V L P Y W T A R T P V E T A R V
963 GCGGAACTGGTCTGCTGCCGTACTGGACAGCCCGCACCCCGTCCGAGACCGCCCGATC
      sp1
```

# IDENTIFICAZIONE DELLA REGIONE 5'- DEL GENE *thr1*




# Thr1

Dominio di adenilazione

 Regioni del promotore

 Possibili codoni d'inizio

 Shine - Dalgarno

 Motivi conservati nei domini di adenilazione

 Primer utilizzati nella TAIL-PCR

```
1  R L G R M R R Q R R P A A I - A A W P L
3  CGACTGGGACGCATGAGGCGACAGCGGCGACCGGCCATTGAGCGGCCTGGCCCCTT

21  L L S G P F L V N S Q R I H R L R F A V
63  TTGCTAAGTGGGCCGTTTCTCGTCAACTCCCAGCGGATTACAGACTCCGATTGCTGTG

41  K T - V Q P E C P N P - R S H V T K I T
123 AAGACATAGGTTGAGCCAGAGTGTCCAAACCCTTGACGATCACATGTAACTAAGATCACA

61  F I Y G G V R T D I D R V T M - C D F G
183 TTTATCTACGGCGGGTCCGGACAGACATCGATCGTGTAACGATGTAGTGCCTGACTTCGGG

81  P R W T G G F G H L H N G A I V S E K K
243 CCACGTTGGACTGGTGGATTCCGGTCATCTACACAATGGAGCCATTGTGTCCGAAAAGAAG

101  D Q Y P S P G Y S E F G A T L V D L F S
303 GACCAGTACCCCTCGCCAGGATATTCGGAATTCGGCGCGACACTGGTTGACTTGTTCCTCC

121  R A A M E M P D R T A L H I D D E K I S
363 AGGGCAGCCATGGAAATGCCCGACCGCACGGCGCTCCACATCGACGACGAGAAGATCTCG
      A1
141  Y G L I H S W A E G L A D L L H D A G V
423 TACGGCCTCCTCCATTCTGGGCGGAGGGCCTCGCGGATCTTCTCCACGACGCCGGGGTA

161  R K G D R V A L R M P P G A N A I A A M
483 CGGAAGGGTGATCGTGTGCGACTGCGTATGCCGCGGGGCGGAACGCGATCGCCGCCATG
      A2
181  L G I L R A G A A Y V P L D I R N P P A
543 CTGGGAATCCTCCGGGCCGGCGCCGCTATGTTCCGCTCGACATTGCAACCCACCCGCT

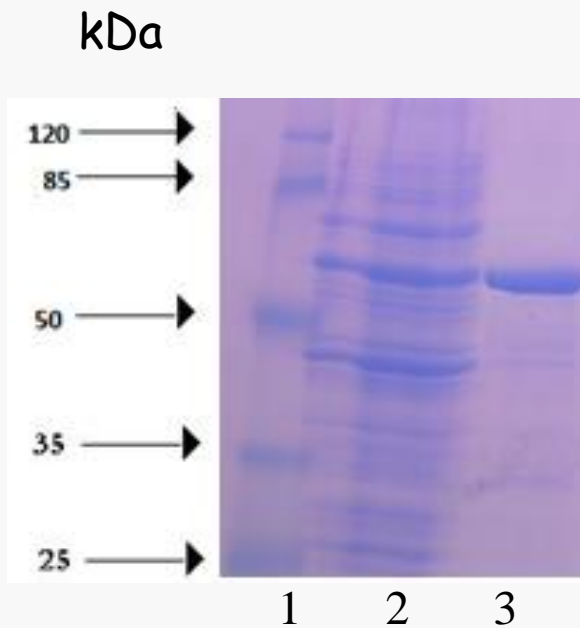
201  R N A F I V T D S Q V V A L V G D P G D
603 CGCAACGCCTTCATCGTCCCGACAGCCAGGTGGTCCGCGTGGTGGGAGATCCCGGTGAT

221  I P E Y T G P L V T E E N V A A L R D R
663 ATTCCCGAATACAGGGGCCGCTCGTCCAGGAGAACGTGGCAGCACTGCGGGACCGT

241  E A I K D P G P E R P G P Q D V A Y I I
723 GAGGCGATCAAGGATCCCGGGCCCGAGCGGCCCGCAGGACGTGCGCTACATCATC
      A3
261  Y T S G T T G R P K G V P V R H G N V T
783 TACACCTCCGGCACACGGGCCGCCCAAAGGGTCCCCGTCGGGCACGGGAATGTCACC
      sp3
281  A L F E A C S R L F S F S A D D R W L L
843 GCTCTCTTCGAGGCCTGCTCCCGCTGTTCTCCTTCTCCGCGGACGACCGCTGGCTGCTC
      A4 sp2
301  F H S M A F D F S V W E I W G A L S T G
903 TTCCACTCCATGGCCTTCGACTTCTCGGTCTGGGAGATCTGGGGTGCCTGAGCACCGGC
      sp1
321  A E L V V L P Y W T A R T P V E T A R V
963 GCGGAACTGGTCGTCCCTGCGTACTGGACAGCCCGCACCCCGTCCGAGACCGCCCGATC
      sp1
```

# STUDI SULLA PROTEINA Thr1

SDS-PAGE della proteina His-Thr1 purificata mediante cromatografia di affinità su Ni-NTA Agarosio

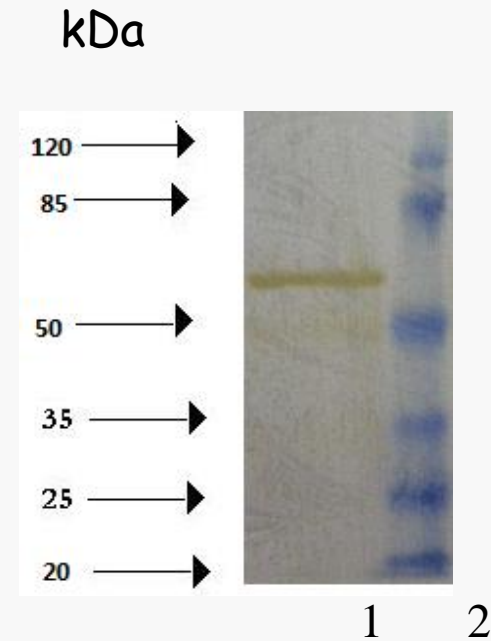


linea 1: marcatori di peso molecolare

linea 2: lisato

linea 3: eluato

Western blotting della proteina His-Thr1 purificata mediante cromatografia di affinità su Ni-NTA Agarosio



linea 1: eluato

linea 2: marcatori di peso molecolare

# STRATEGIE DI MUTAGENESI

## Mutagenesi sito-specifica

Per valutare il ruolo strutturale/funzionale di specifici aminoacidi, si producono poche varianti sulla base di ipotesi predefinite

## Mutagenesi casuale (evoluzione guidata in vitro)

Per modificare le proprietà di una proteina (termostabilità, specificità di substrato), si produce un gran numero di varianti che devono essere analizzate per identificare quelle migliori

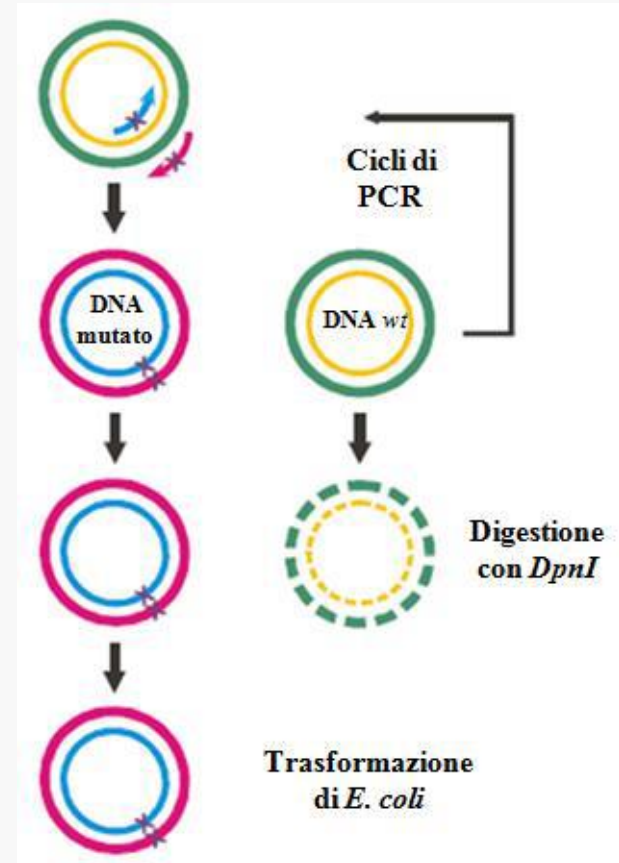
PCR error-prone

DNA shuffling

saturation

## Mutagenesi sito-specifica QuikChange

1. PCR di amplificazione dell'intero plasmide con due oligo complementari che portano la mutazione
2. Digestione con DpnI per degradare il template wild type (metilato)
3. Trasformazione di *E. coli*
4. Verifica mediante sequenziamento del DNA



## Mutagenesi sito-specifica: introduzione simultanea di mutazioni multiple (overlap extension PCR)

- PCR per amplificare diversi frammenti contenenti le mutazioni A B C parzialmente sovrapposti
- I frammenti amplificati vengono mescolati e sottoposti a due cicli di denaturazione ed annealing
- PCR con oligonucleotidi esterni per permettere di completare l'estensione dei filamenti mancanti

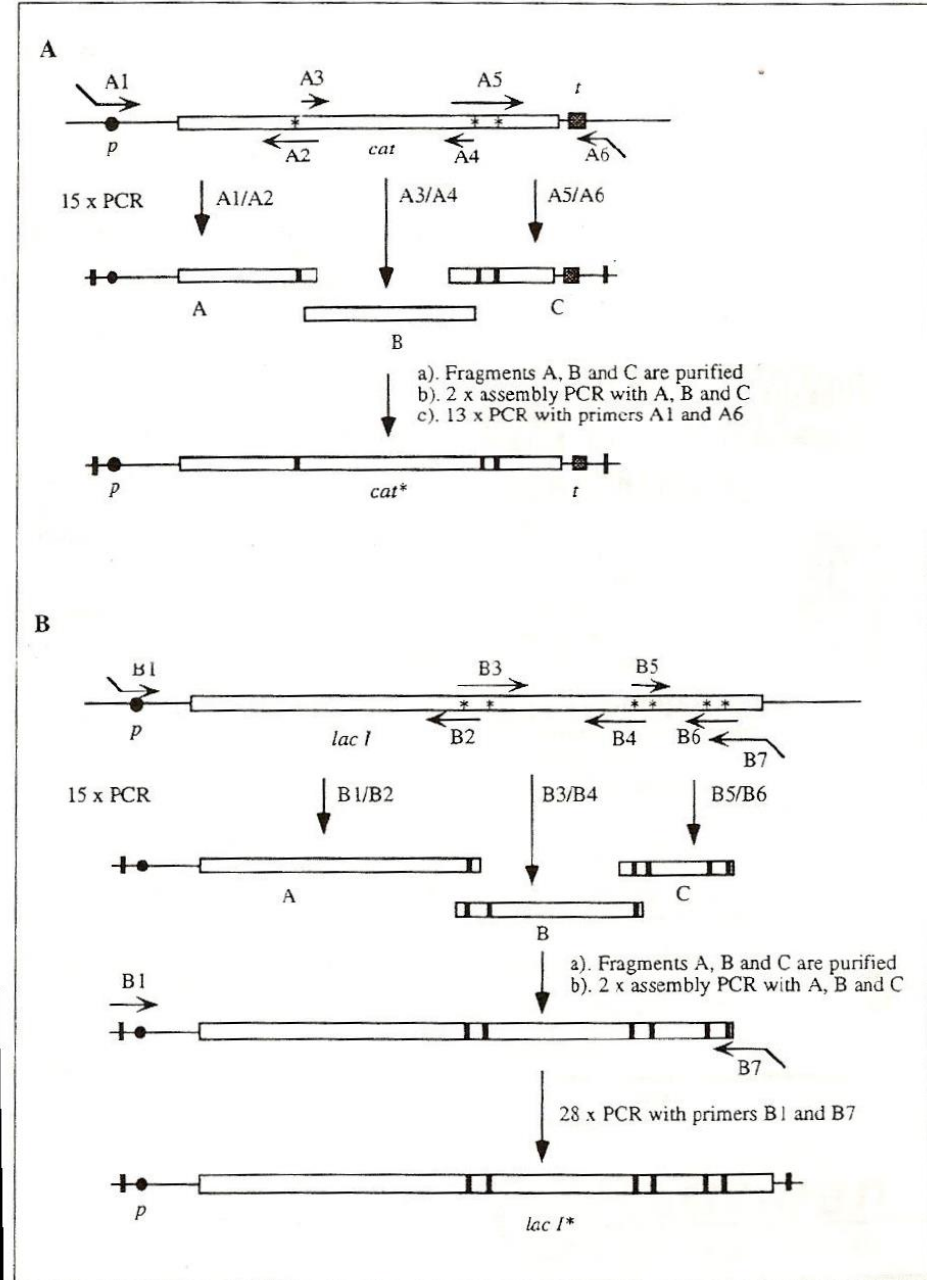


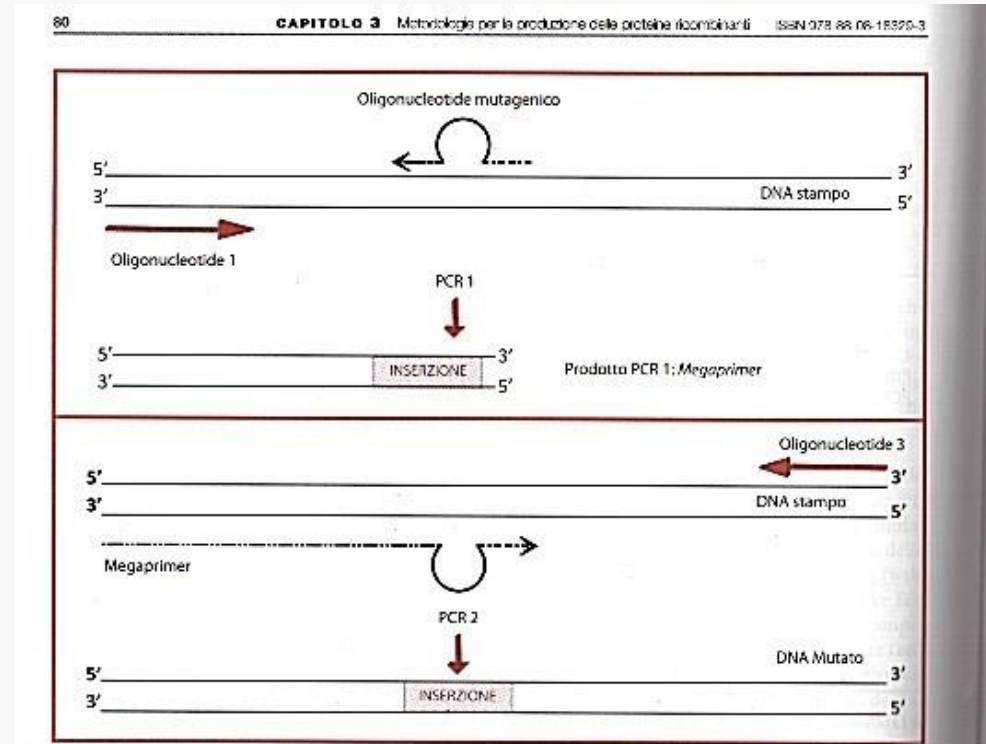
Figure 1. Outline of the simultaneous multisite mutagenesis procedure. The *cat* (A) and *lacI* (B) gene segments were amplified, purified and assembled to produce complete genes with introduced mutations. *P*, promoter; *t*, terminator; \*, restriction site that is to be changed; and solid bar, introduced mutation.



## Mutagenesi sito-specifica: introduzione simultanea di mutazioni multiple (megaprimer)

Si eseguono 2PCR consecutive con due diverse coppie di oligo  
1PCR oligo esterno e oligo contenente l'inserzione fiancheggiata da almeno 10 basi per l'appaiamento allo stampo a monte e a valle

2PCR con frammento ottenuto che funzionerà da megaprimer e oligo 3 esterno





## Mutagenesi casuale

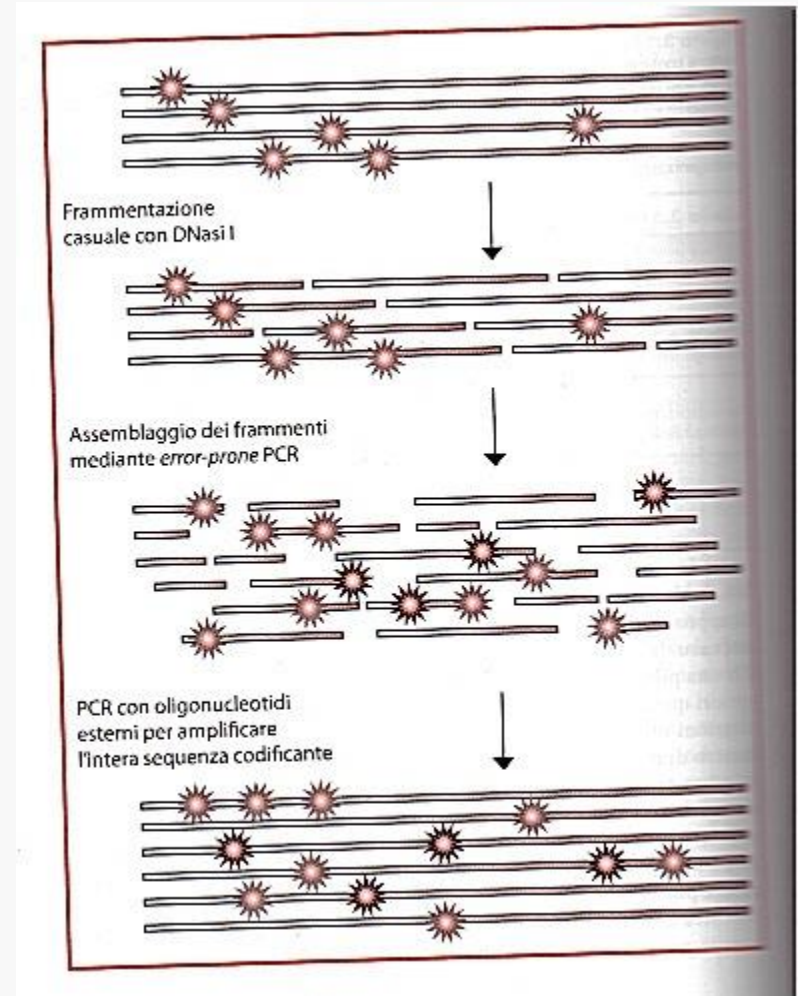
### Evoluzione guidata in vitro: error-prone PCR

- PCR in condizioni tali da aumentare la frequenza di errore della Taq polimerasi:
  - aggiungendo  $Mn^{2+}$  (riduce la specificità di appaiamento delle basi),
  - sbilanciando le concentrazioni dei desossiribonucleotidi (se uno dei dNTP diventa limitante, potrebbe essere incorporato al posto un altro dNTP ),
  - aumentando la concentrazione di  $Mg^{2+}$  (stabilizza l'appaiamento di basi non complementari),
  - aumentando la concentrazione di Taq polimerasi (aumenta la probabilità di allungamento di estremità che derivano da inneschi non corretti)

## Mutagenesi casuale Evoluzione guidata in vitro: DNA shuffling

Il DNA proveniente da geni omologhi  
) o da esperimenti di mutagenesi casuale  
viene

- Digerito con DNAsi o enzimi di restrizione
- Riassembleto in una PCR senza oligo
- Amplificato con oligo esterni per ricostituire un'intera sequenza codificante ibrida



## SCREENING

PER IDENTIFICARE I MUTANTI CON LE PROPRIETA' DESIDERATE

# Sistemi di espressione

## Procariotici

**E. Coli, B. Subtilis**

## Eucariotici:

**Saccharomyces Cerevisiae (lievito)**

**Cellule di insetto (Baculovirus)**

**Cellule di mammifero in coltura (CHO etc.)**

**Animali transgenici**

**Piante transgeniche**

# Scelta del sistema di espressione

Si valuta la qualità del prodotto finale:

Modifiche post.sintetiche necessarie per il corretto folding

Attività biologica della proteina

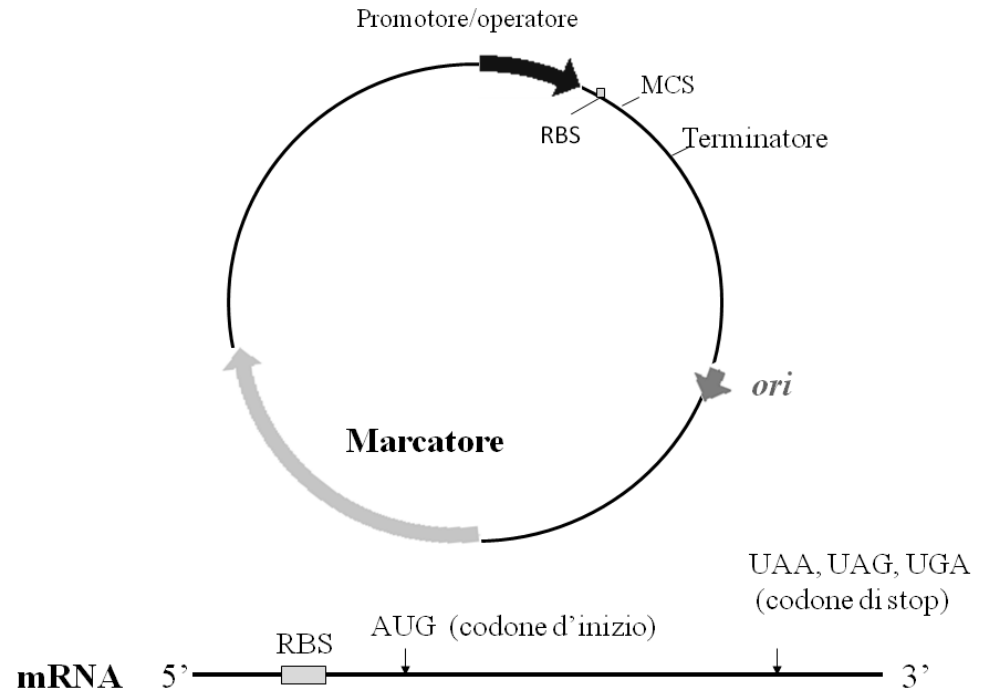
## Principali modifiche post-sintetiche delle proteine

- Formazione dei ponti disolfuro
- Glicosilazione
- Processamento proteolitico
- $\gamma$ -carbossilazione,  $\beta$ -idrossilazione
- Fosforilazione
- Acetilazione, metilazione, ADP-ribosilazione, amidazione
- Solfatazione

## Vettori di espressione

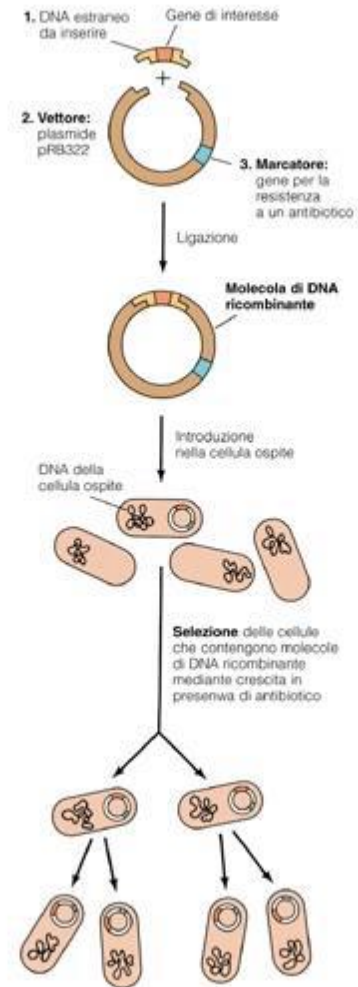
### Elementi specifici per l'ospite:

- Elementi di regolazione della trascrizione (promotore e terminatore)
- Marcatore di selezione
- Origine di replicazione (assente nei vettori integrativi)



# Escherichia Coli

- Crescita rapida (ca 30 min/raddoppio)
- Sistema economico
- Caratterizzazione dettagliata dei meccanismi molecolari dei processi di trascrizione e traduzione
- Ceppi mutanti
- Trasformazione semplice (CaCl<sub>2</sub>, elettroporazione)
- Elevati livelli di espressione da plasmidi multicopia con promotori forti inducibili
- Selezione mediante resistenza ad antibiotici
- Espressione preferenzialmente intracellulare



# Vettori di espressione procariotici

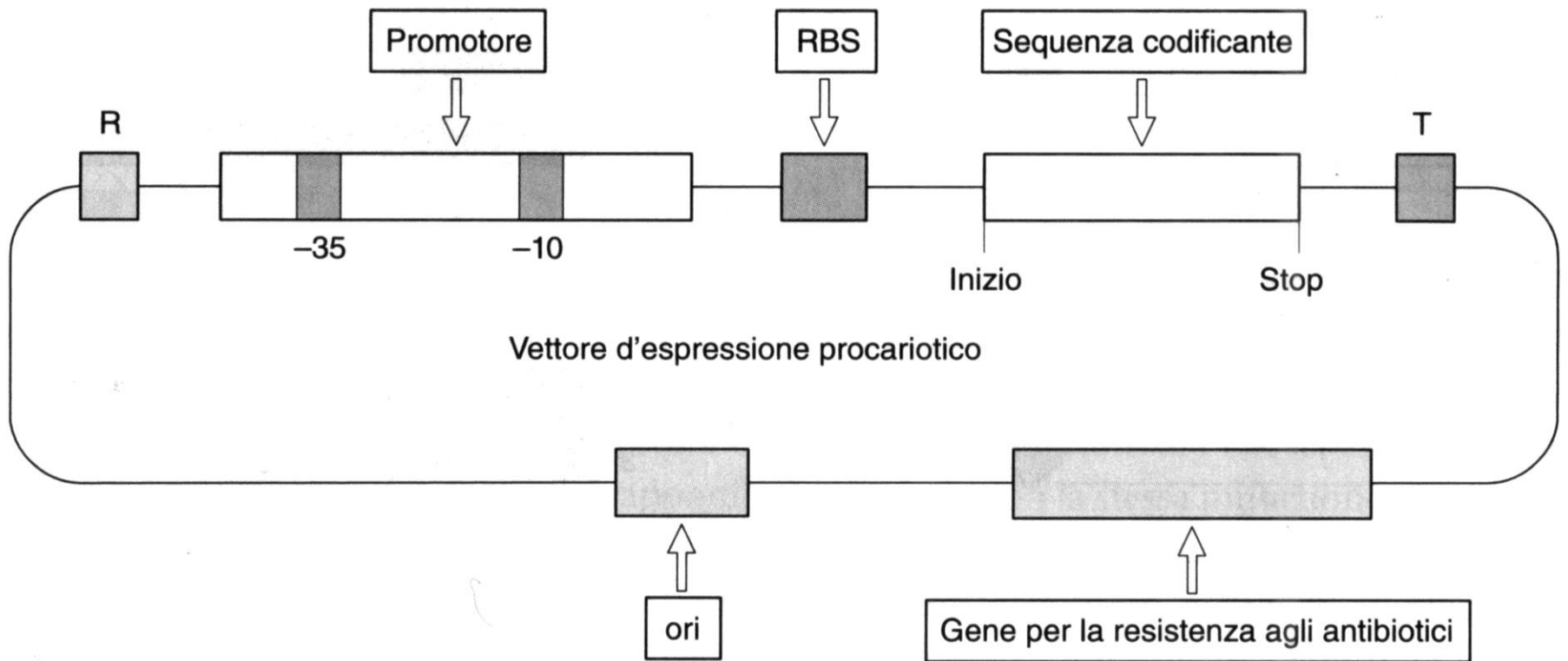


Figura 3.34 Componenti di un tipico vettore d'espressione. Per produrre un trascritto (sequenza codificante) e tradurlo, il vettore deve possedere alcune sequenze, tra cui il promotore e il sito di legame ai ribosomi (RBS). L'attività del promotore può essere modulata da un gene regolatore (R), che agisce in modo simile al gene regolatore dell'operone *lac*. Il terminatore della trascrizione è indicato da T.



## PROMOTORI FORTI E REGOLABILI

Forti

– Alta affinità per l'RNA polimerasi  TRASCritto ALTA FREQUENZA

Regolabile

– L'espressione può essere controllata dallo sperimentatore, tramite induttori e repressori

### UNA PRODUZIONE CONTINUA DELLA PROTEINA PROVOCA:

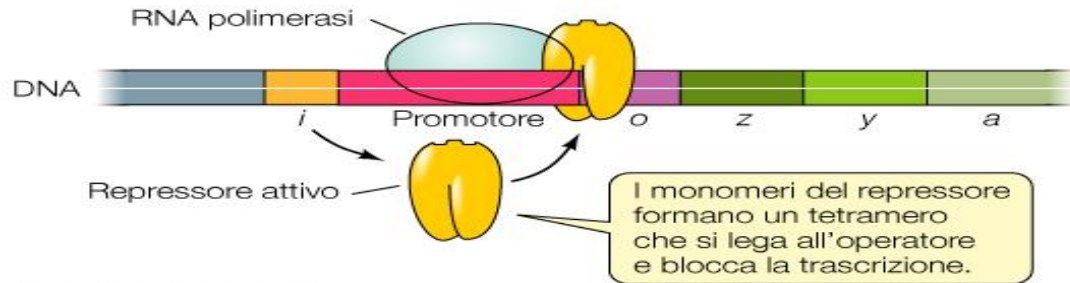
- inibizione funzioni cellula
- perdita energia
- perdita plasmide

## **E. coli: promotori forti e inducibili**

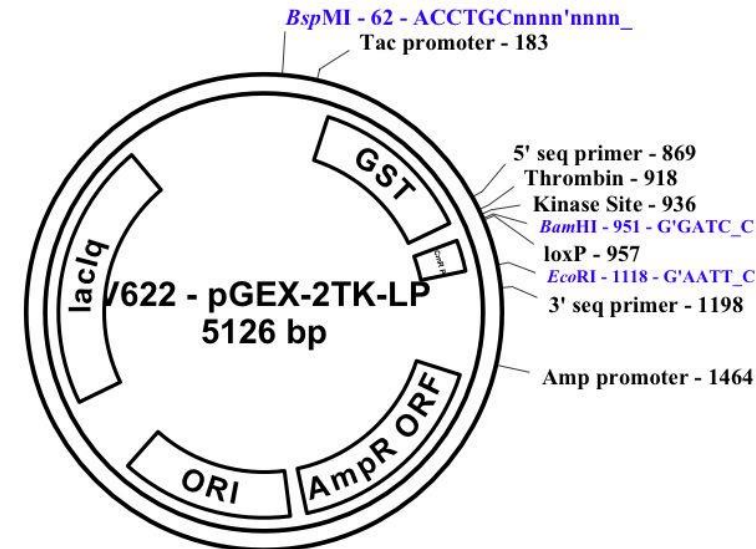
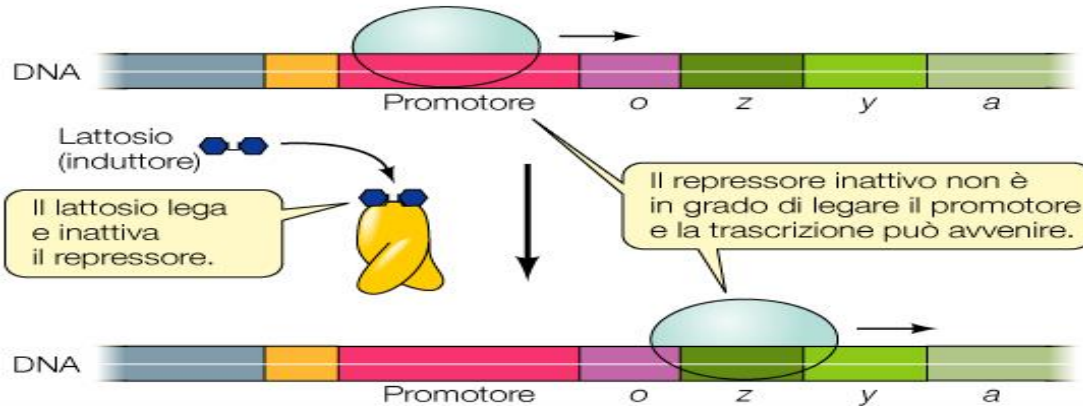
- Promotore lac (induzione con IPTG, un analogo del galattosio)
- Promotore trp (induzione in assenza di triptofano)
- Promotore tac: regione -35 promotore trp -10 promotore lac
- Promotore costitutivo fago T7 (deve essere presente nel ceppo la RNA polimerasi di T7)

# Promotore lac

## Assenza di lattosio

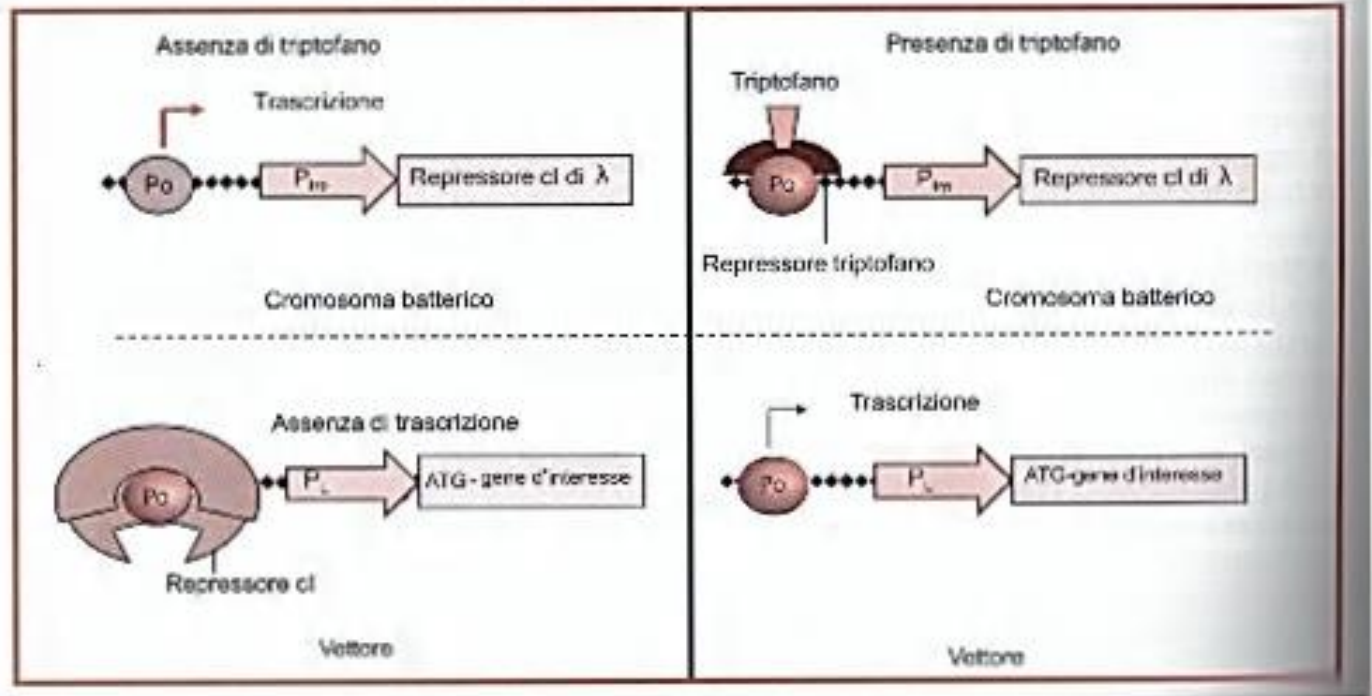


## Presenza di lattosio



# Promotore trp

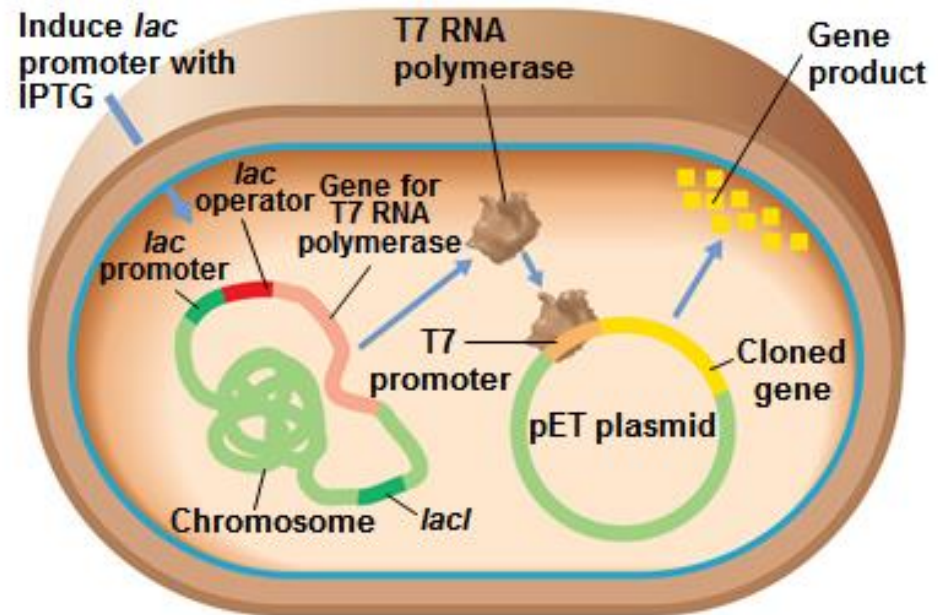
mostrato in figura 3.4.



# Promotore T7

Figure 11.17

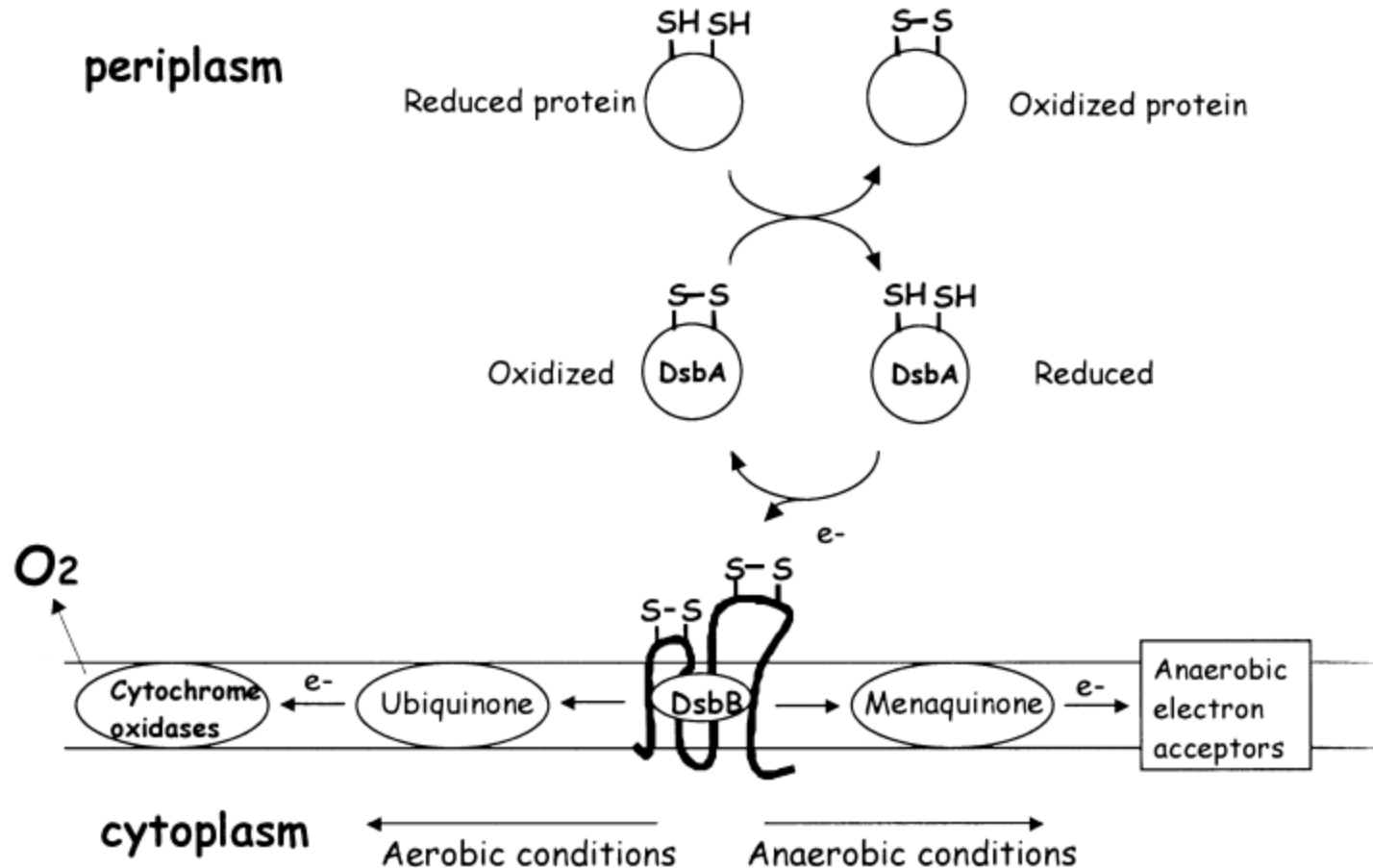
Il gene della RNA polimerasi T7 è integrata nel cromosoma del ceppo ospite sotto il controllo del promotore lac. L'espressione dell'RNA polimerasi è indotta da IPTG



# Esporto della Proteina in E.Coli

- L'ambiente ossidante del periplasma permette la formazione dei ponti disolfuro, cosa che non avviene nell'ambiente riducente del citoplasma.
- Il periplasma contiene la disolfuro ossidoreduttasi (DsbA) e la disolfuro isomerasi (DsbC), che catalizzano la formazione e l'isomerizzazione dei ponti disolfuro.
- La stabilità della proteina può aumentare (meno proteasi).
- Si permette l'accumulo di proteine che possono essere tossiche nel citoplasma
- La proteina può essere isolata più facilmente
- N-terminale vero
- **Svantaggio:** minori livelli di espressione

# Formazione ponti disolfuro in E.coli



# PROTEINE RICOMBINANTI IN PROCARIOTI

## Vantaggi

- sistemi di espressione molto semplici da manipolare
- produzione di proteine in grandi quantità ed a basso costo

## Problemi

E' difficile ottenere proteine eucariotiche che mantengano la conformazione nativa dato che i batteri mancano di adeguati sistemi di modificazione post-traduzionale (glicosilazione, tagli proteolitici, fosforilazione, aggiunta di lipidi).



# Saccharomyces cerevisiae

- Crescita rapida (ca 90 min/raddoppio)
- Sistema economico
- Caratterizzazione dettagliata dei meccanismi molecolari dei processi di trascrizione e traduzione
- Ceppi mutanti (auxotrofie,)
- Trasformazione semplice (LiCl, elettroporazione)
- Plasmidi episomali o integrativi con promotori forti inducibili (GAL1, PHO5) o costitutivi (ADH, GAPD, TPI, PGK)
- Selezione mediante complementazione di una auxotrofia o resistenza ad antibiotici (G418)
- Espressione intracellulare o secreta (sequenza segnale a-factor)
- Modifiche post-sintetiche

I vettori di espressione eucariotici hanno le stesse caratteristiche dei vettori procariotici

Promotore (p)

Sito di clonaggio (cs)

Segnali di arresto e poliadenilazione (t)

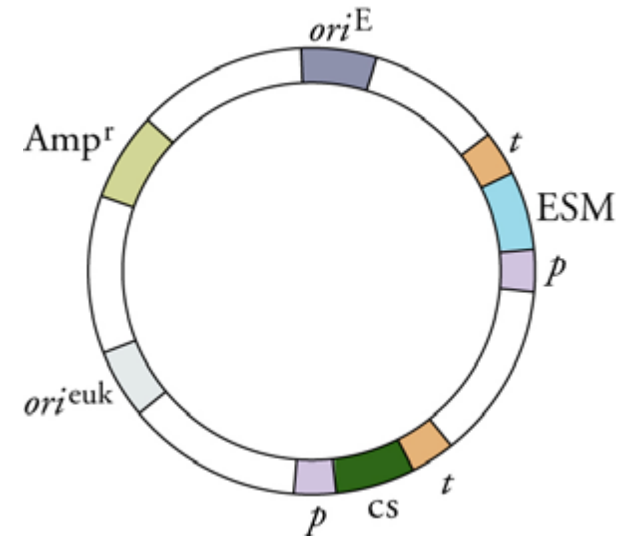
Origine della replicazione eucariota (orieuk)

Gene marcatore selezionabile eucariota (ESM)

Origine della replicazione procariota (oriE)

Gene marcatore selezionabile procariota (Amp<sup>r</sup>)

per permettere la manipolazione in E. coli



## Promotori in Saccharomyces

Acid phosphatase **PHO5** Inducibile

Alcohol dehydrogenase I **ADH I** Costitutivo

Alcohol dehydrogenase II **ADH II** Inducibile

Galactokinase **GAL 1** Inducibile

Metallothionein **Cup 1** Inducibile

Phosphoglycerate kinase **PGK** Costitutivo

Triose Phosphate isomerase **TPI** Costitutivo

1. **Vettori episomiali o plasmidici YeP (Yeast episomal plasmid)**

Molto diffusi

Copie multiple

Possibile instabilità su larga scala (>10 L)

2. **Vettori che si integrano YiP Yeast integrating plasmid**

Limitazioni del numero di copie del gene clonato (uno per cromosoma)

e quindi limitazione della resa in proteina

Si integrano per ricombinazione omologa

3. **Cromosomi artificiali di lievito (YAC)**

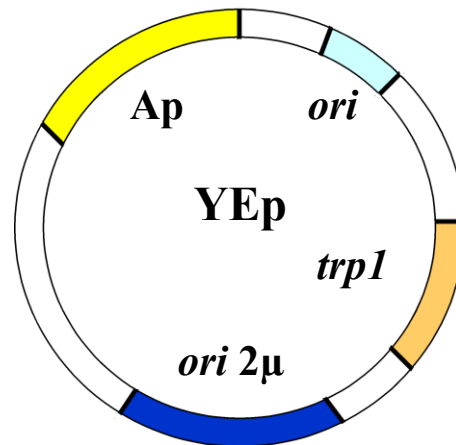
Lungo frammento di DNA che viene mantenuto nella cellula come cromosoma separato

Altamente stabile

Non ancora usati per produrre proteine eterologhe su scala industriale

# *Saccharomyces cerevisiae*

I plasmidi episomiali, chiamati **YEp** sono basati su un plasmide endogeno di lievito di tipo episomale, il plasmide 2 micron. Contengono un'origine per la replicazione autonoma ( $2\ \mu$ ). Più copie del plasmide (50-100 per cellula) sono propagate nelle cellule di lievito e mantenute come episomi.

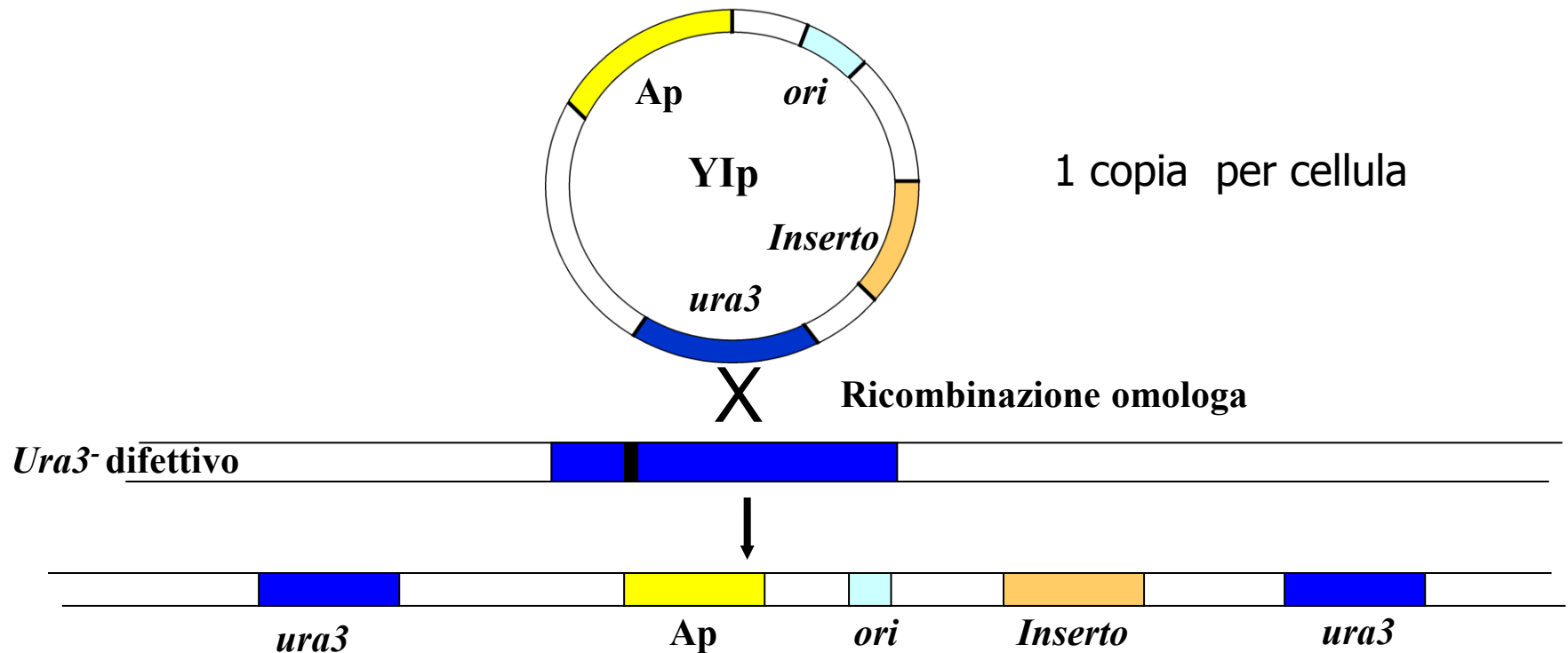


# Saccharomyces cerevisiae

**Yip** vettore di integrazione, deriva da un vettore plasmidico batterico.

- marker di selezione di lievito
- manca di origine di replicazione

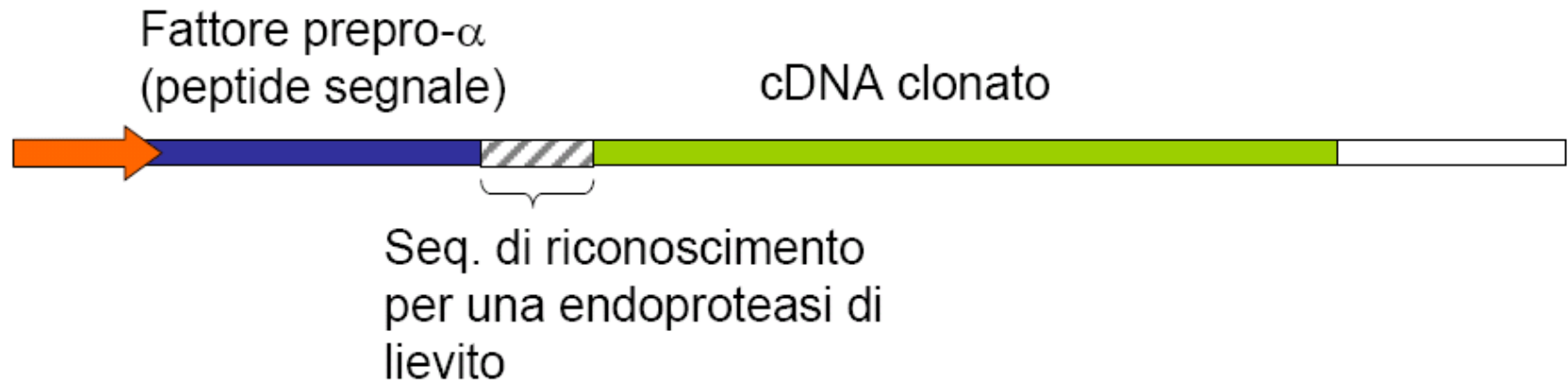
Integrazione per ricombinazione omologa tra un marcatore funzionale di lievito e il corrispondente marcatore difettivo in un lievito mutante. L'integrazione porta alla duplicazione della sequenza bersaglio. Il DNA duplicato fiancheggia il vettore. I plasmidi integrativi sono propagati stabilmente, ma occasionalmente possono essere "spinti fuori" per ricombinazione tra le due sequenze duplicate.



# *Saccharomyces cerevisiae*

## Secrezione

Si possono utilizzare peptidi segnale di lievito, oppure si può clonare il cDNA di interesse contenente la sequenza codificante il peptide segnale.



Durante il processo di esportazione si verificano:

1. la formazione di ponti disolfuro,
  2. il taglio proteolitico
  3. le modificazioni post-traduzionali
- ...e la proteina si libera nel mezzo extracellulare

## Problemi nella espressione di proteine ricombinanti in *S. cerevisiae*:

- Perdita del plasmide (anche con promotori inducibili)
- Iperglicosilazione della proteina eterologa [fino a 100 residui di mannosio contro ~10 normale]=> alterazione della attività biologica e/o immunogenicità
- A volte proteine progettate per essere secrete sono trattenute nello spazio periplasmatico



## *Pichia pastoris*

- Lievito metilotrofico
- Crescite ad alta densità
- Glicosilazione più simile a quella di mammifero
- Promotori inducibili **AOX1** (promotore dell'alcol ossidasi indotto da metanolo), **CUP1** (rame)
- Promotori costitutivi **GAPD** (gliceraldeide fosfato deidrogenasi), **YPT1** (pro
- Sistemi di esporto delle proteine più efficienti Limitata secrezione nel mezzo di proteine endogene (quella eterologa arriva a rappresentare il 30% del secreto
- Possibilità di “umanizzare” il pathway glicosilativo (GlycoFi)

AOX: ossidazione di MeOH a formaldeide e H<sub>2</sub>O<sub>2</sub>( CAT).

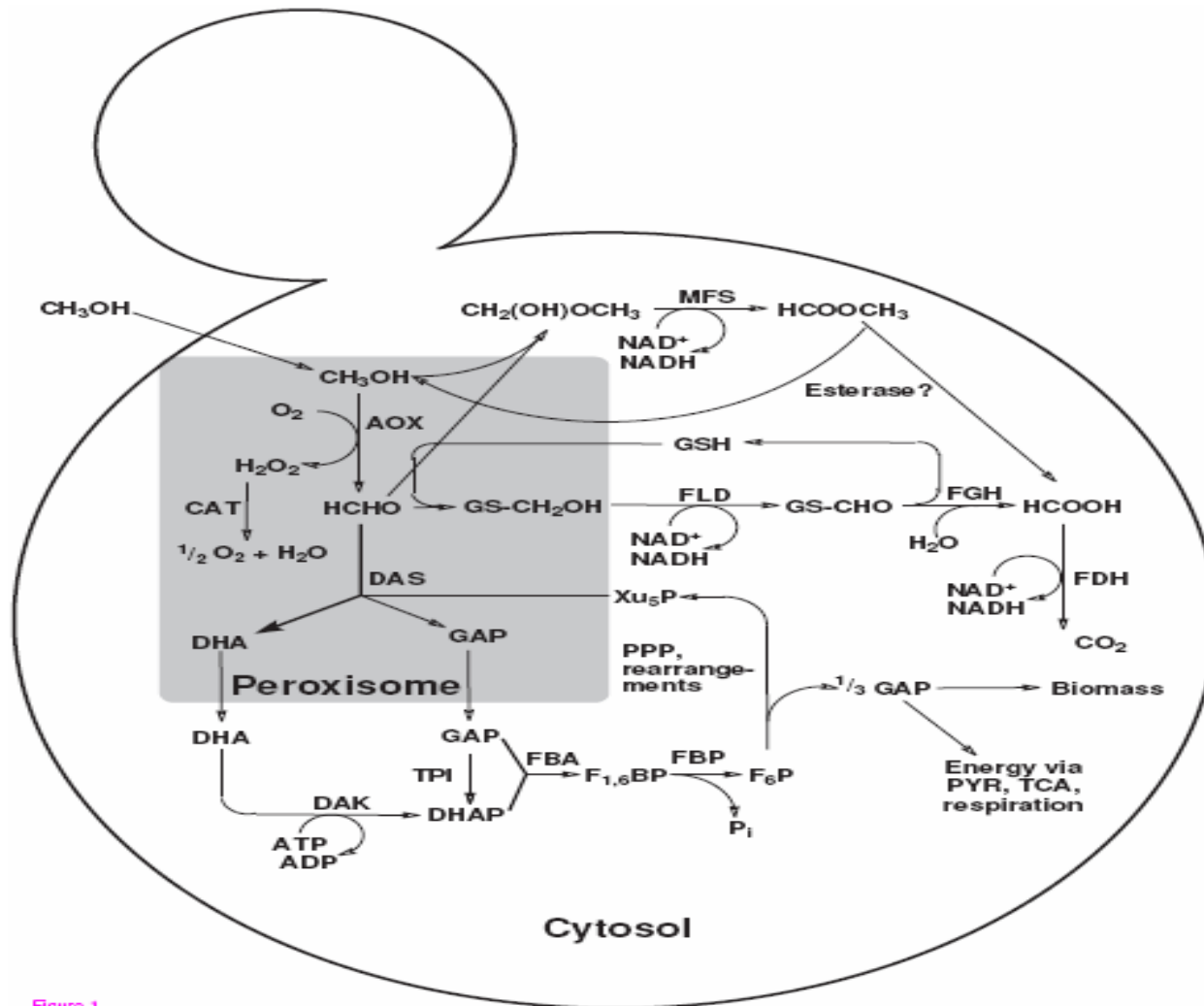
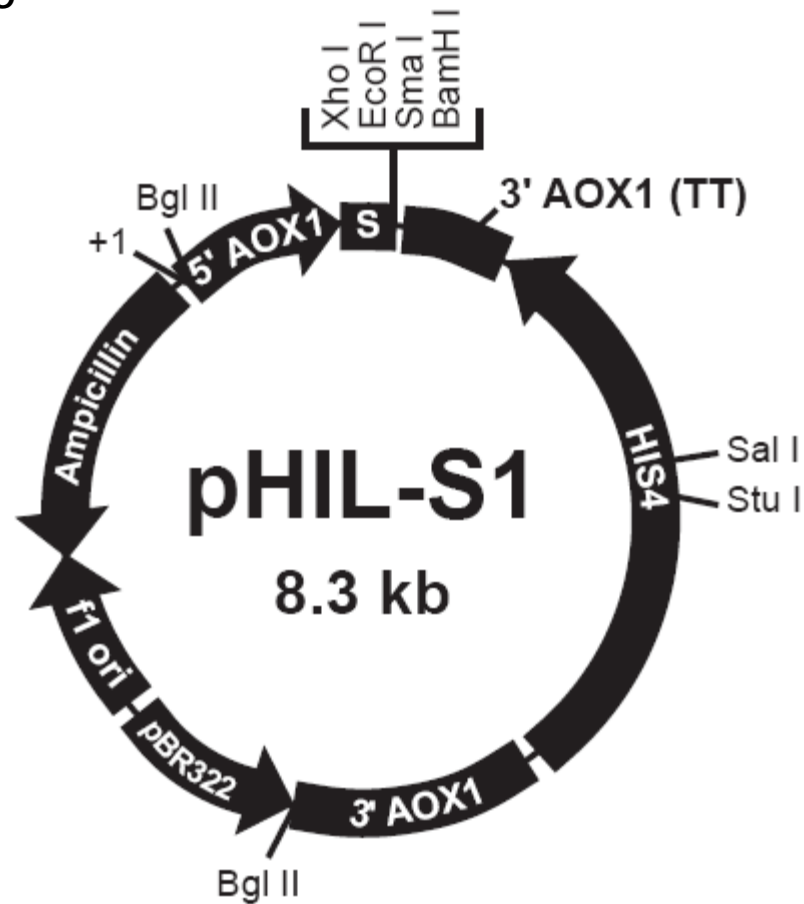
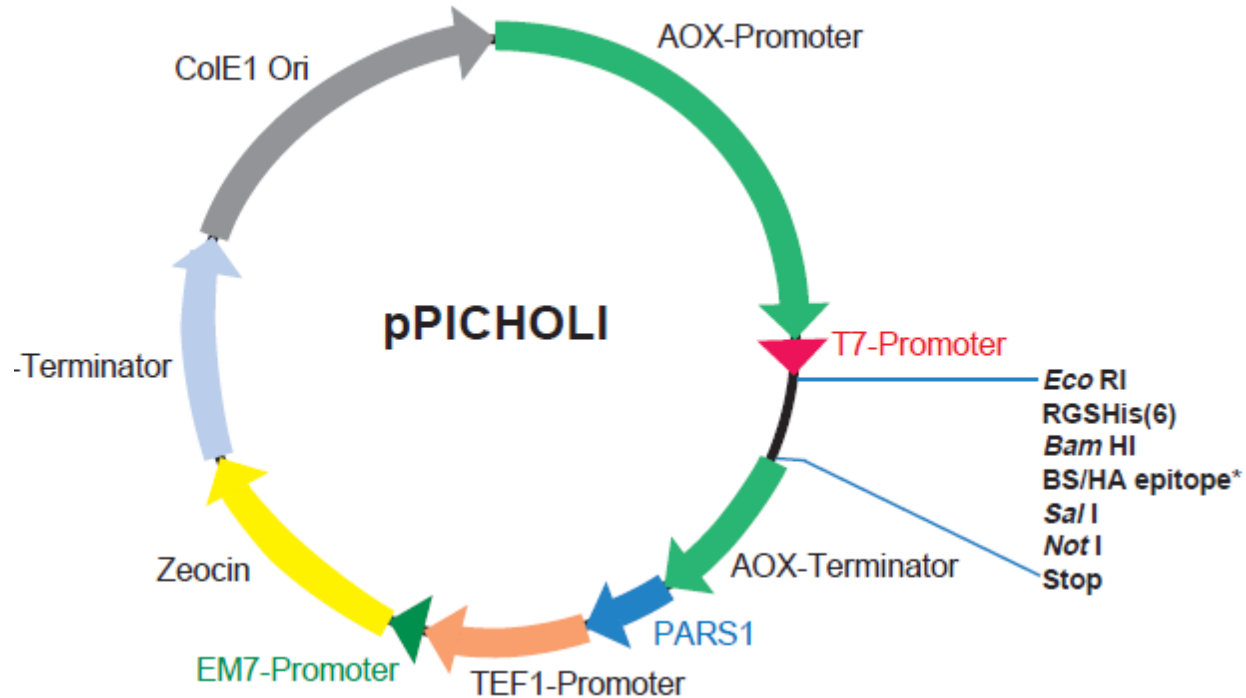


Figure 1

# Vettore integrativo



Vettore shuttle permette l'espressione della proteina sia in *E. coli* (pT7) che in *Pichia Pastoris* (pAOX1)



## Cellule di mammifero

- Linee cellulari derivate da tumori CHO (Chinese Hamster Ovary), BHK (Baby Hamster Kidney) , HEK293(Human Embryonic Kidney)
- Sistemi molto complessi e costosi
- Crescita lenta (ca 24 ore/raddoppio)
- **Plasmidi episomali o integrativi** con promotori forti inducibili (fitoestrogeni, TetOn/TetOff) o costitutivi (citomegalovirus CMV)
- Trasfezione transiente o stabile (lipocomplessi, elettroporazione)
- Selezione mediante resistenza ad antibiotici o vantaggio selettivo (diidrofolato reductasi, glutamina sintasi)
- Adatte alla produzione di proteine terapeutiche folding e modifiche post-sintetiche generalmente corrette

## Vettori di espressione eucarioti

- **Origine di replicazione**

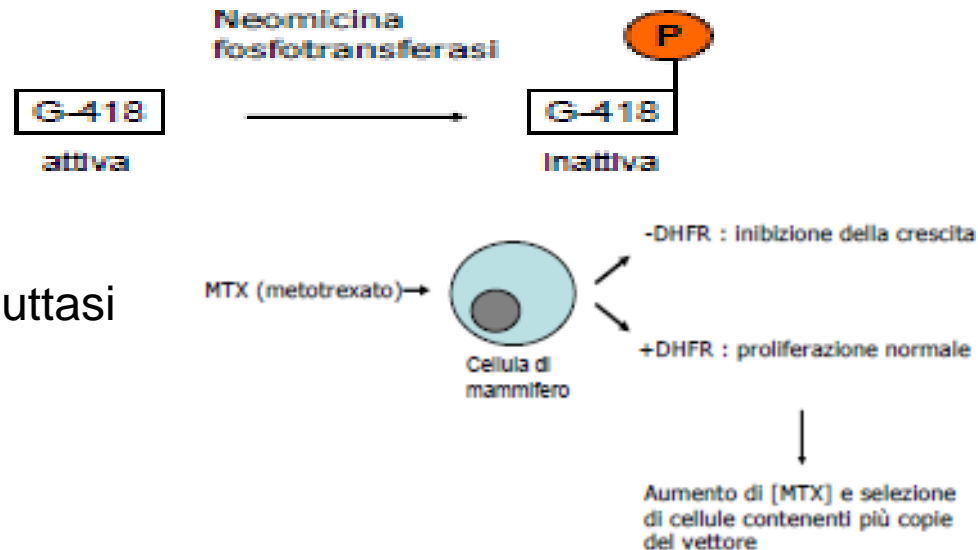
- Generalmente derivata da virus animali (SV40 Simian virus 40)

- **Promotori/sequenze di arresto**

- Generalmente derivati da virus animali o da geni di mammifero altamente espressi SV40, cytomegalovirus (CMV), herpes simplex virus (HSV) actina  $\beta$ , timidina chinasi, ormone della crescita bovino

- **Marcatori selezionabili**

**Neo:** neomicina fosfotransferasi fosforila G-418 inattivandolo



- **DHFR** deidrofolato reduttasi

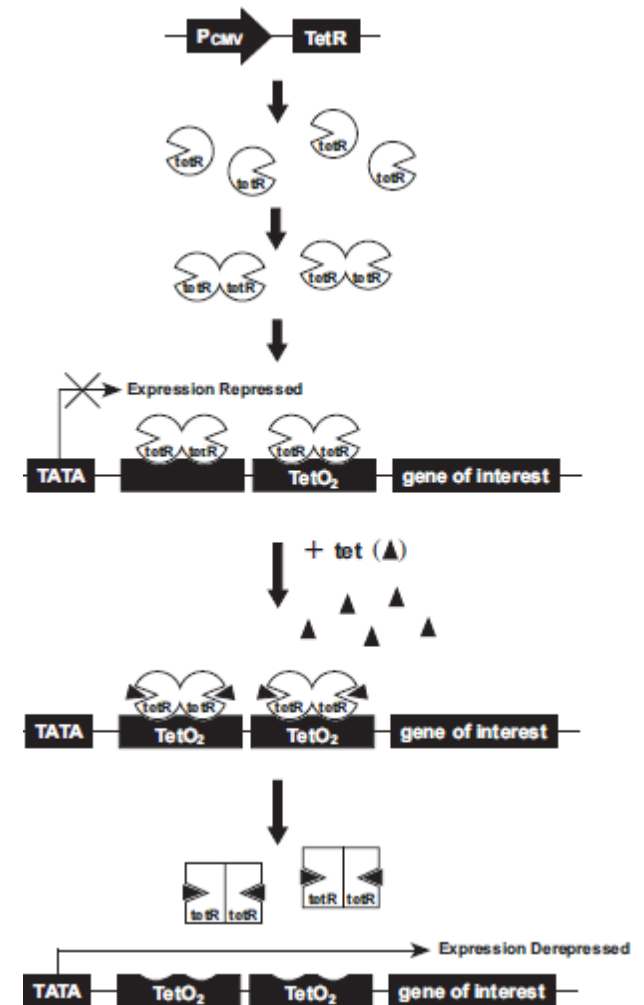
## Induzione dell'espressione indotta da tetraciclina in cellule di mammifero

### Sistema TetOn per l'espressione regolata da tetraciclina

Il repressore batterico TetR lega sequenze specifiche di DNA in assenza di tetraciclina

Il gene di interesse è posto sotto il controllo di un promotore ibrido tetO-CMV e viene inserito in una linea cellulare che esprime costitutivamente TetR

In presenza di tetraciclina TetR non si lega al DNA e permette la trascrizione del gene di interesse

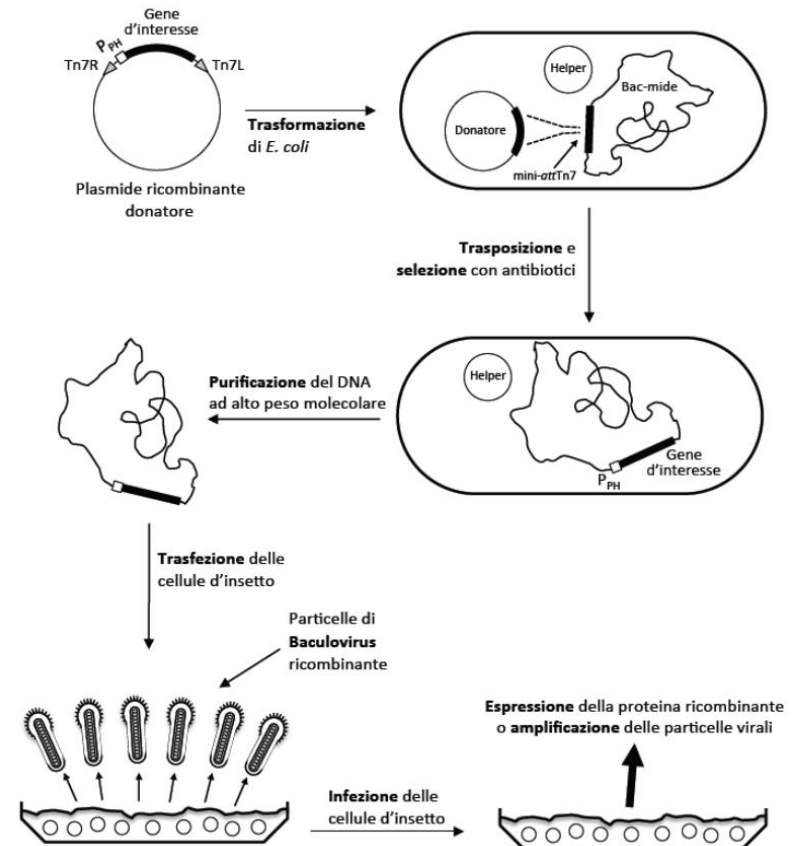


# Cellule di insetto/baculovirus

I baculovirus formano una famiglia di virus il cui aspetto somiglia a un bastoncino i Baculoviridae. Sono distinti in due sottofamiglie: Eubaculovirinae, con Nucleopoliedrovirus (NPV) e Nudibaculoviridae Granulovirus (GV), e. Questi virus possono infettare soprattutto specie di insetti come le larve delle tarme e dei sinfiti, le zanzare e anche i crostacei. Il genoma dei baculovirus è costituito da un DNA circolare a doppia catena, di dimensione compresa tra 80 e 180 kbp, che ne fa un virus di grossa taglia. I baculovirus sono impiegati in biotecnologia per la produzione di proteine ricombinanti dal 1983. Questi virus permettono di introdurre il gene codificante della proteina da produrre nelle colture di cellule di insetti, in particolare nelle cellule di *Spodoptera frugiperda* (Sf9 e Sf21) e di *Trichoplusia ni* (Hi5). Il cDNA della proteina in questione viene integrato nel genoma del virus, sostituendo il gene codificante alla proteina di struttura fibrillare p10, c. Un esempio di proteina prodotta in baculovirus è la proteina **interferone**

## Cellule Sf9 infettate con baculovirus ricombinante

- Promotore forte della poliedrina
- Crescita lenta (ca 24 ore/raddoppio)
- Sistemi molto complessi e costosi
- Folding e modifiche post-sintetiche generalmente corrette





## Miglioramento dei livelli di espressione di una proteina ricombinante

- **Miglioramento della traduzione**

- Sequenze di legame per i ribosomi
- Stabilità dell'mRNA
- Codon usage

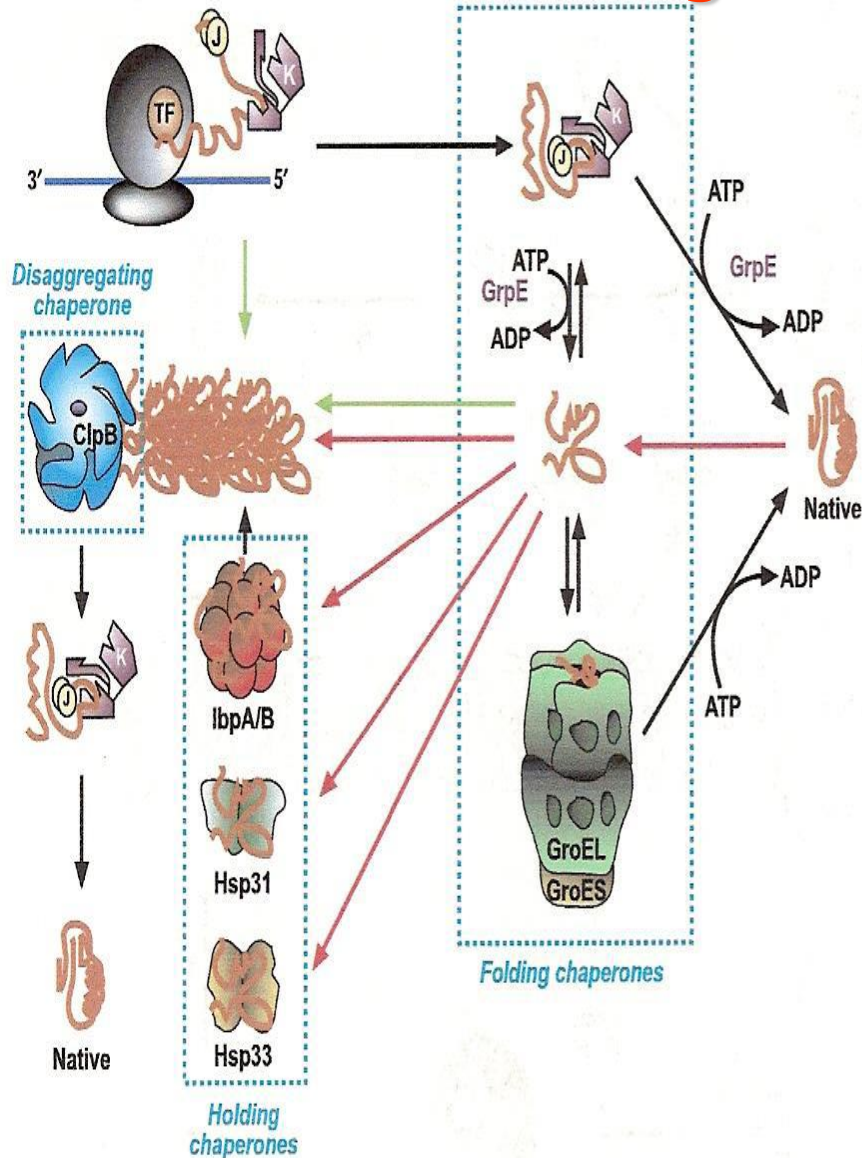
- **Miglioramento del folding**

- Variazione delle condizioni di crescita (temperatura, composizione del terreno)
- Coespressione del macchinario di folding (chaperon molecolari)
- Proteine di fusione

- **Miglioramento della stabilità**

- Secrezione (è necessario un segnale di secrezione all'N-terminale)
- Proteine di fusione
- Modifiche post-traduzionali

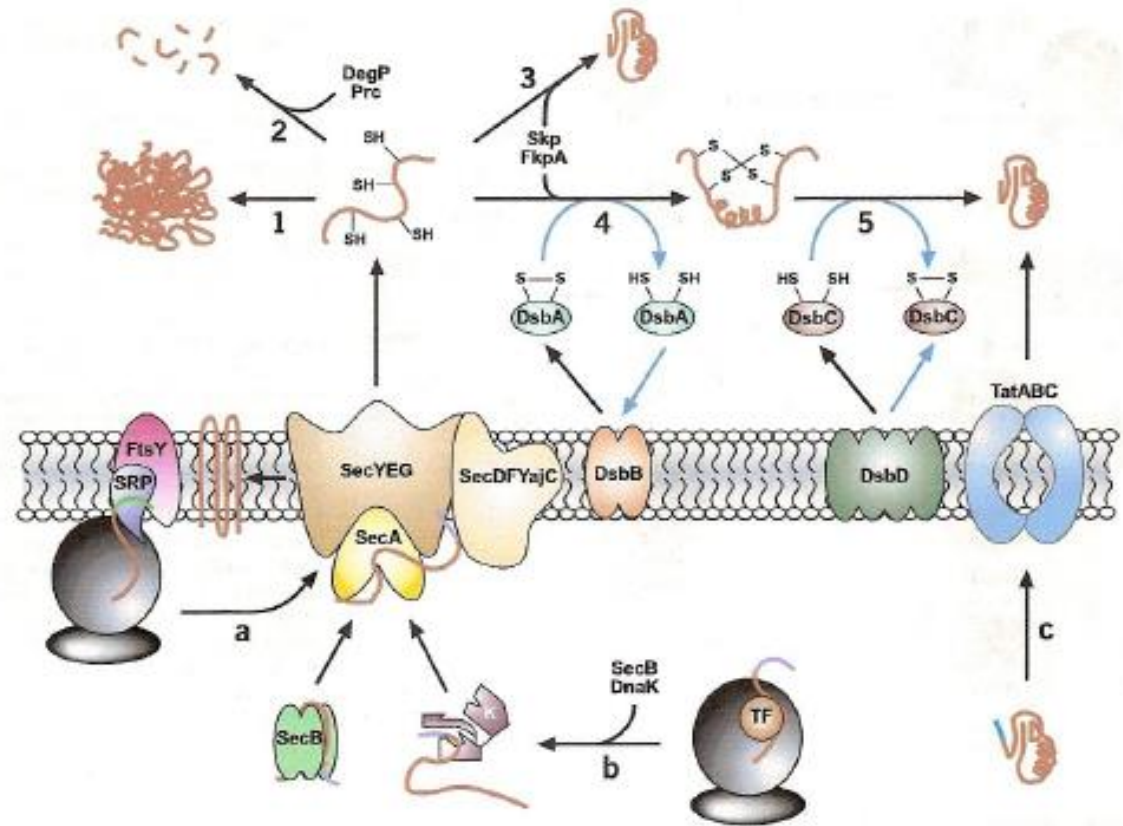
# Folding delle Proteine in E.Coli



**Figure 1** Chaperone-assisted protein folding in the cytoplasm of *E. coli*. Nascent polypeptides requiring the assistance of molecular chaperones first encounter TF or DnaK-DnaJ. Both chaperones engage solvent-exposed stretches of hydrophobic amino acids, shielding them from the solvent and each other. After undocking from TF- or GrpE-mediated release from DnaK, folding intermediate may reach a native conformation, cycle back to DnaK-DnaJ or be transferred to the central chamber of GroEL for folding at infinite dilution upon GroES capping. In times of stress (red arrows), thermolabile proteins unfold and aggregate. IbpB binds partially folded proteins on its surface to serve as a reservoir of unfolded intermediates until folding chaperones become available and intercalates within large aggregates. The holding chaperones Hsp33 and Hsp31 become important under oxidative and severe thermal stress, respectively. ClpB promotes the shearing and disaggregation of thermally unfolded host proteins and cooperates with DnaK-DnaJ-GrpE to reactivate them once stress has abated. Recombinant proteins that miss an early interaction with TF or DnaK/DnaJ, that undergo multiple cycles of abortive interactions with folding chaperones or titrate them out, accumulate in inclusion bodies (green arrows).

# Meccanismi di folding e secrezione in *E. coli*

- **Peptide segnale** fortemente idrofobico viene riconosciuto da SRP il complesso SRP ribosoma lega il recettore FtsY e poi trasferito al traslocone SecYEG e SecDFYajC
- **Peptide segnale** mediamente idrofobico viene intercettato SecB che lega la catena distesa e poi SecA e SecYEG
- **Peptide segnale** contenente 2 arginine viene traslocato attraverso il sistema TatABC dopo che ha raggiunto la forma nativa

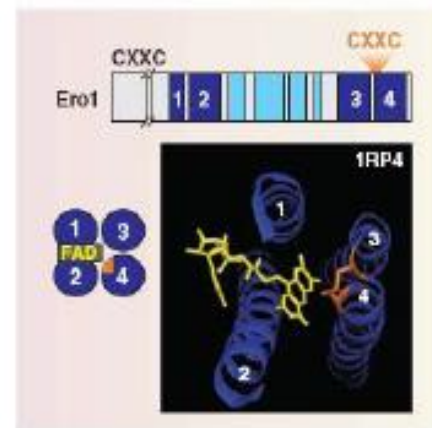
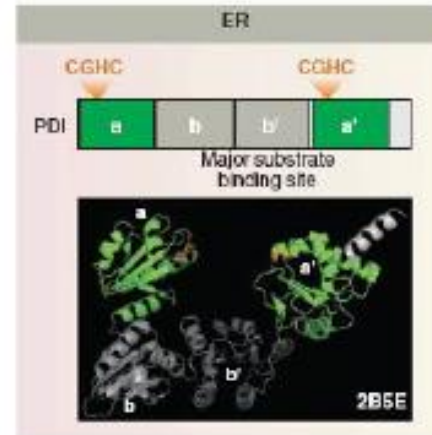
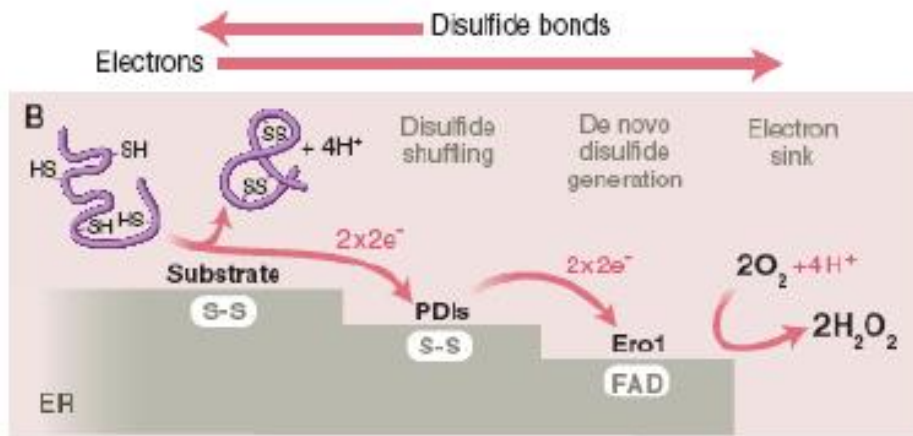


**Figure 2** Export and periplasmic folding pathways. Proteins destined for export can be translocated across the inner membrane in three different fashions. (a) Preproteins with highly nonpolar signal sequences (green) or transmembrane segments of inner membrane proteins are recognized by SRP which, along with TF, scans nascent chains. SRP-dependent export involves delivery of the ribosome-nascent chain complex to FtsY and subsequent translocation through the SecYEG-SecDFYajC translocon. (b) The vast majority of preproteins have less hydrophobic signal sequence (lavender) and undergo Sec-dependent export. TF associates with the nascent polypeptide, halting cotranslational folding. As the chain grows, TF dissociates and the polypeptide is transferred to SecB or DnaK that maintain it in an extended conformation. Delivery to SecA and ATP-dependent translocation through SecYEG completes the process. (c) Preproteins with signal sequences containing the twin-arginine motif (cyan) are exported via the Tat-dependent pathway in a folded form. After cleavage of the signal sequence, partially folded periplasmic proteins may aggregate (1), undergo proteolysis (2) or reach a native conformation, possibly with the assistance of folding modulators (3). Cysteine pairs in proteins containing disulfide bonds are oxidized by DsbA (4) whereas incorrect disulfides are isomerized by DsbC (5). These oxidoreductases are reactivated by DsbB and DsbD, respectively. Black arrows show products obtained after each step, whereas blue arrows represent electron flow.

# Meccanismi di folding e secrezione in cellule eucariotiche

- Reticolo endoplasmatico
  - Rimozione peptide segnale
  - Inizio glicosilazione
  - Formazione ponti disolfuro e folding
- Apparato di Golgi
  - Maturazione oligosaccaridi
- Vescicole di secrezione
  - Trasporto alla membrana plasmatica e secrezione

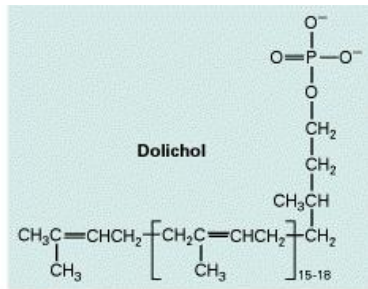
Formazione dei ponti disolfuro nel reticolo endoplasmatico  
 PDI: proteina disolfuro isomerasi  
 Ero1: sulfidril-ossidasi





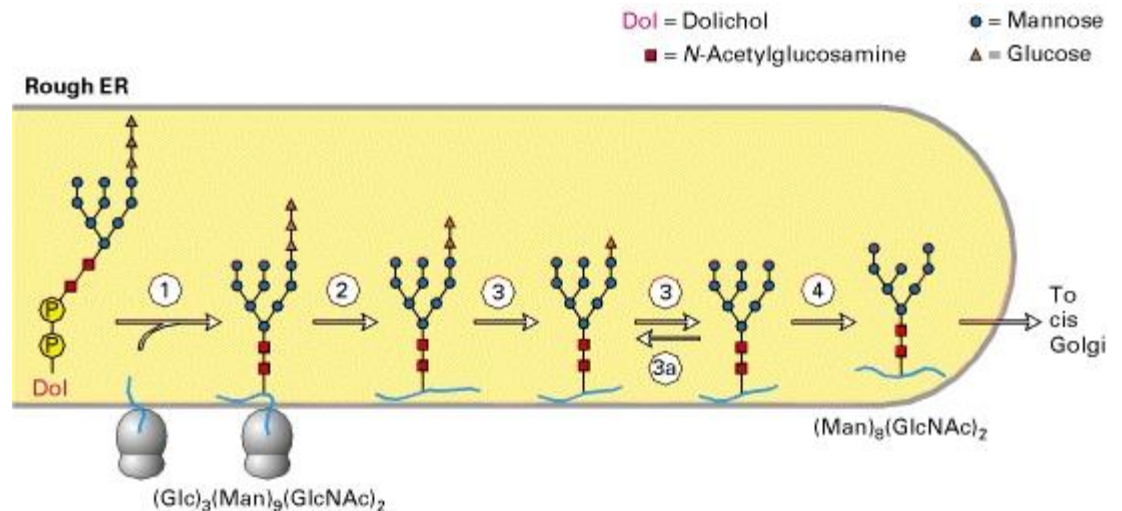
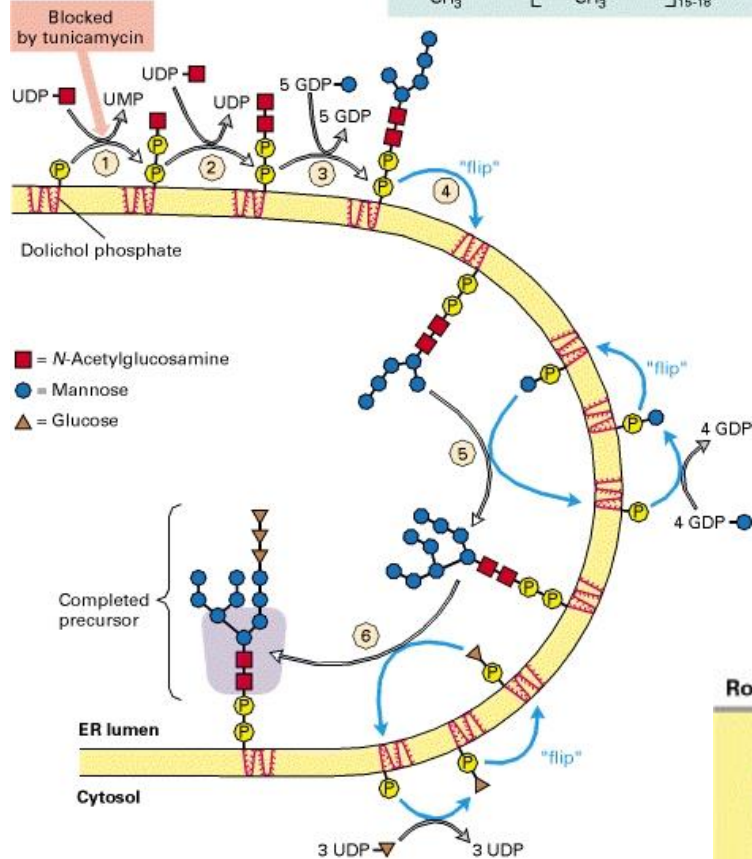
# Glicosilazione delle proteine. Ruolo della glicocomponente

- La glicosilazione può assistere il **folding** della catena polipeptidica attraverso effetti sulla struttura secondaria
- La glicosilazione può indirizzare una proteina alla sua destinazione finale (**trafficking/sorting**)
- Le catene oligosaccaridiche possono contribuire alla **stabilità** di una proteina aumentandone la solubilità, schermando regioni idrofobiche, proteggendo dalle proteasi o favorendo interazioni intra-catena
- Le catene oligosaccaridiche possono partecipare al **riconoscimento/legame** con ligandi o recettori o all' **attività biologica** di una proteina
- L'acido sialico che si trova all'estremità delle catene oligosaccaridiche aumenta l' **emivita** delle proteine plasmatiche
- Le catene oligosaccaridiche possono contribuire all' **immunogenicità** di proteine eterologhe

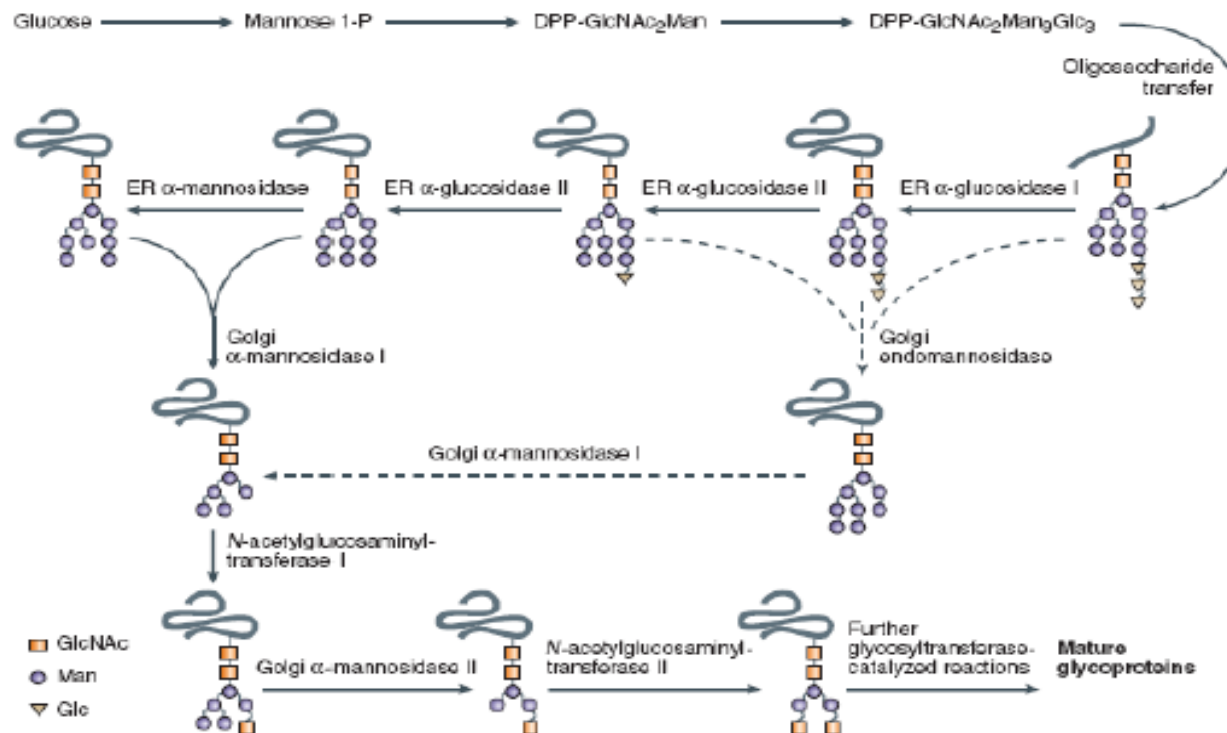


## Nglicosilazione delle proteine

- Formazione del nucleo oligosaccaridico a per successive aggiunte di monosaccaridi sul dolicolo fosfato
- Traslocazione nel lumen del reticolo
- Trasferimento mediato dall'oligosaccariltrasferasi sulla sequenza consenso N-X-T/S



# N-glicosilazione delle proteine in corrispondenza di motivi di sequenza N-X-T/S

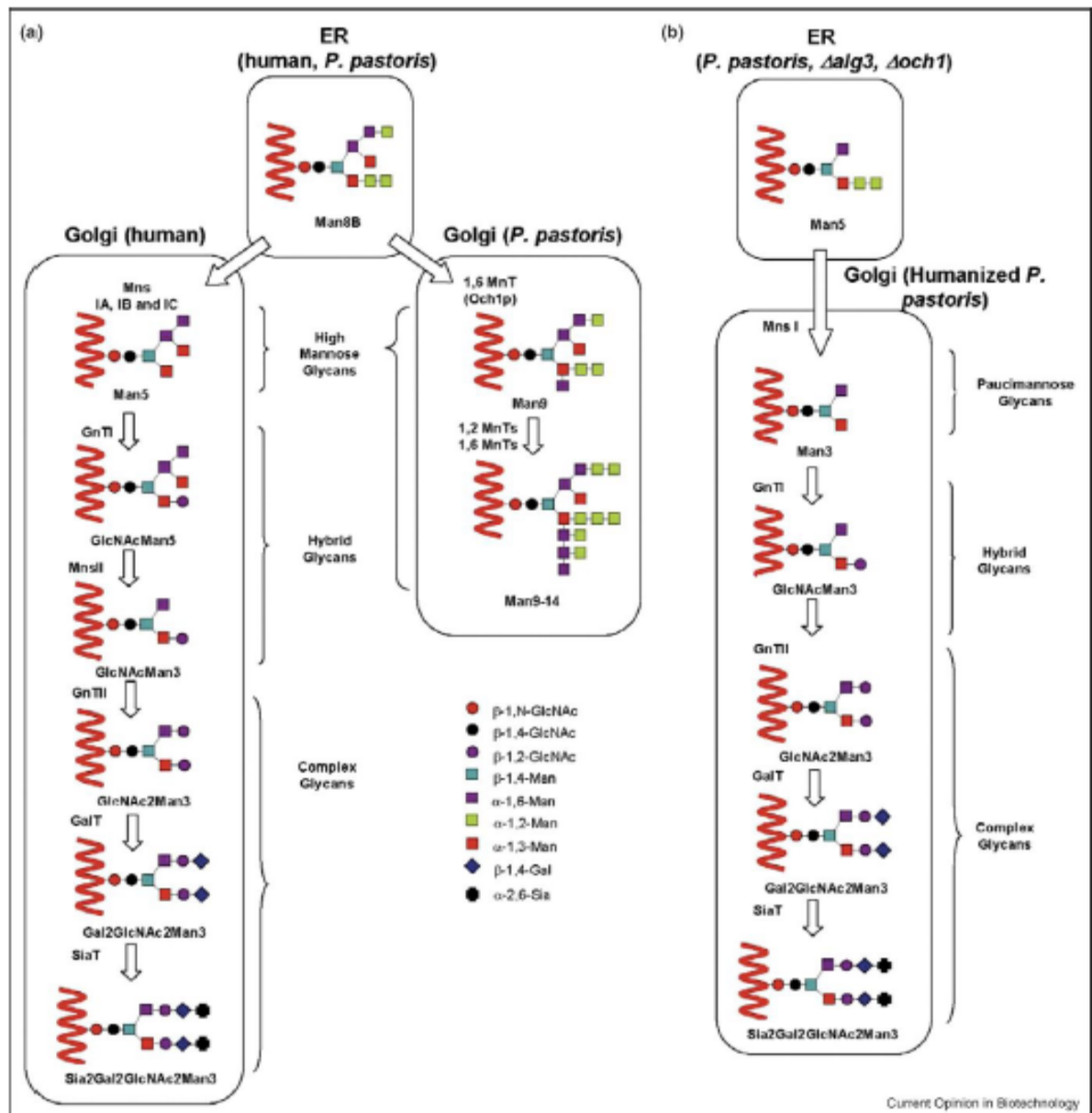


**Figure 1** Overview of the biosynthesis of protein N-linked glycan side chains. As the nascent glycoprotein enters the ER, a preformed oligosaccharide, dolichol phosphate precursor (DPP) is attached cotranslationally to some asparagine residues that are part of the consensus sequence Asn-Xaa-Ser/Thr (where Xaa is any amino acid). The biosynthesis of this precursor, its attachment to asparagine residues and the subsequent steps of its processing in the ER and in the Golgi are performed by a series of glycosidases and glycosyltransferases. The exact profile of such enzymes present in the cell will largely dictate the composition and structure of the final oligosaccharide side chain, which are often described as 'complex', 'high mannose' or 'hybrid' type. Secretion at various stages of the carbohydrate elaboration yields heterogeneous glycoprotein products. (Reprinted by permission from *Nature Reviews Drug Discovery*, Macmillan Magazines Ltd., ref. 136.)

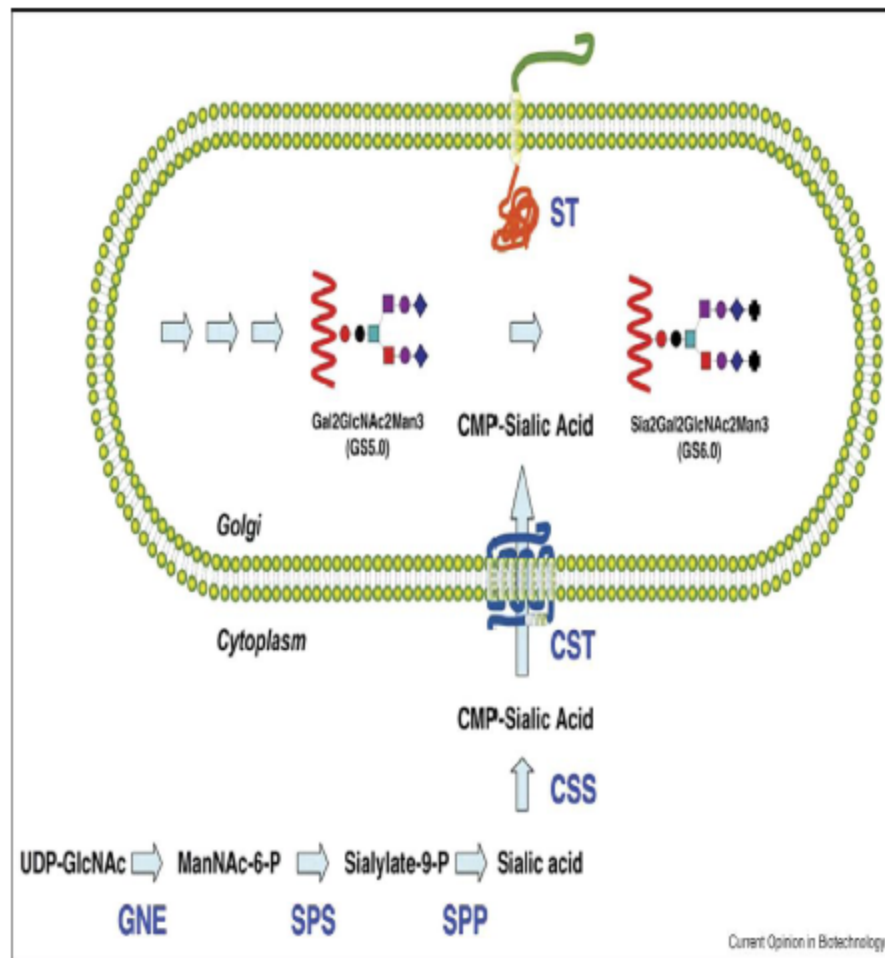


# Umanizzazione della glicosilazione del lievito *Pichia pastoris*

N-linked glycosylation pathways in humans and yeast. Representative N-linked glycosylation pathways in humans and *P. pastoris* (a). An alternative humanized N-linked glycosylation pathway in *P. pastoris* (b). Mns;  $\alpha$  1,2-mannosidase, MnsII; mannosidase II, GnTI;  $\beta$  1,2-N-acetylglucosaminyltransferase I, GnTII;  $\beta$  1,2-N-acetylglucosaminyltransferase II, GalT;  $\beta$  1,4-galactosyltransferase, SiaT;  $\alpha$  2,6-sialyltransferase, MnT; mannosyltransferase. For simplicity the two GlcNAc residues present at the reducing end of all glycans have been omitted.



Glycoengineering steps required for sialic acid transfer in the yeast Golgi. Endogenous UDP-GlcNAc, present in the yeast cytoplasm, is converted to CMP-sialic acid by UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase (GNE), N-acetylneuraminase-9-phosphate synthase (SPS), sialylate-9-P phosphatase (SPP) and CMP-sialic acid synthase (CSS). Subsequently, the product is translocated into the Golgi by the CMP-sialic acid transporter (CST) and sialic acid is transferred onto the acceptor glycan by sialyltransferase (ST). Enzymes are indicated by blue text and metabolic intermediates by black text. For simplicity the two GlcNAc residues present at the reducing end of all glycans have been omitted.



# Purificazione delle proteine ricombinanti

- Metodi convenzionali: frazionamento con solfato d' ammonio, cromatografia a scambio ionico, gel-filtrazione o cromatografia d' interazione idrofobica
- Cromatografia di affinità sfruttando **tag** proteici o peptidici fusi alla proteina (N- o C-terminali). Immobilizzazione di anticorpi contro i tag peptidici o di substrati dei tag proteici

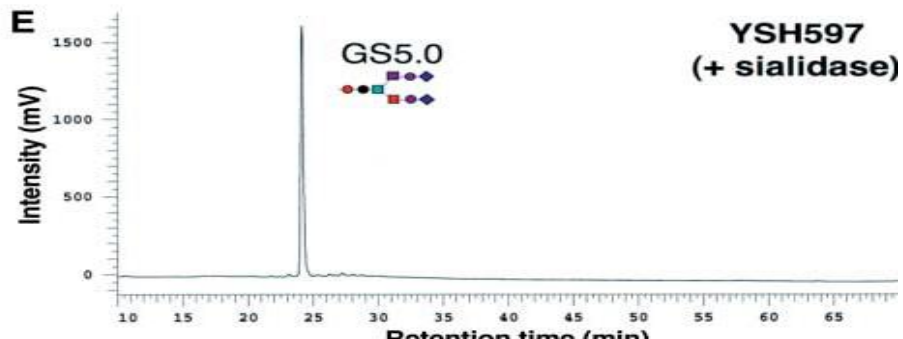
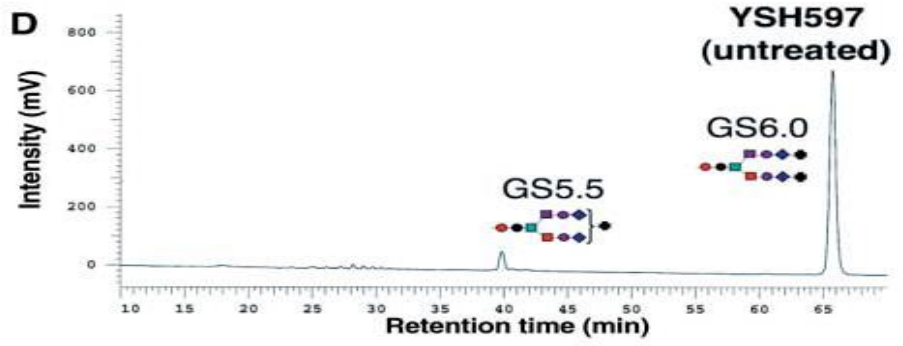
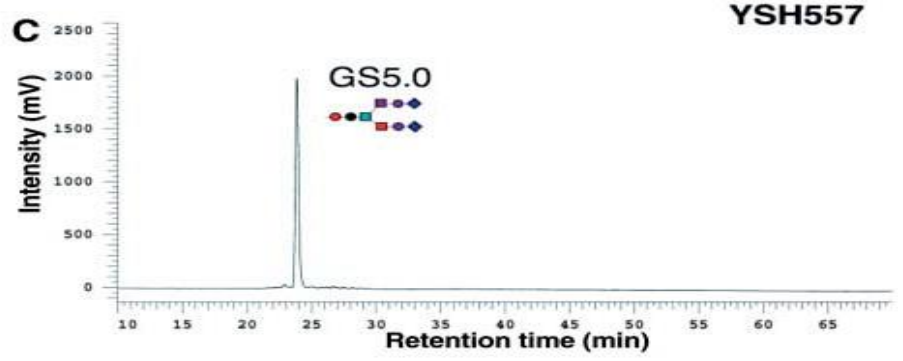
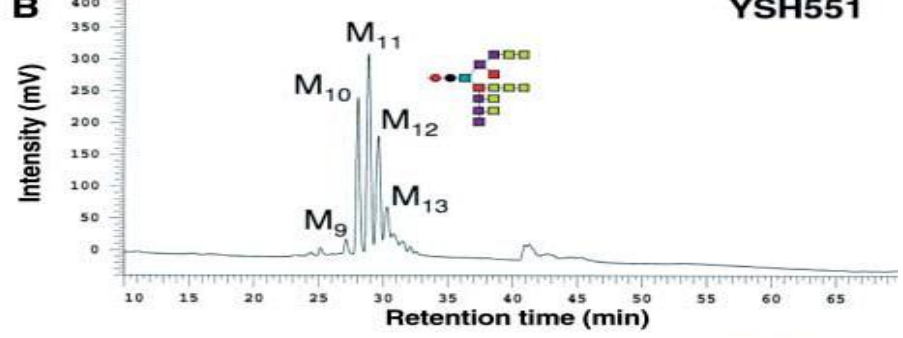
Tag peptidici: 6xHis, flag, c-myc

Tag proteici: GST (glutathione-S-trasferasi), MBP (maltose binding protein), TRX (tioredossina)

Rimozione tag (inserimento siti di taglio per proteasi in una sequenza linker tra la proteina e il tag)

# Produzione di eritropoietina in *P. pastoris*

- glicoproteina necessaria per la maturazione dei globuli rossi con tre siti di Nglicosilazione
- Utilizzata come farmaco per l'anemia



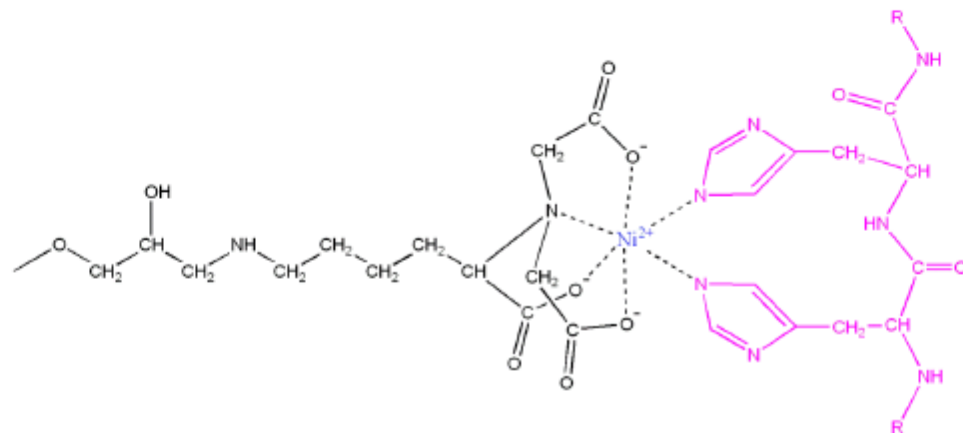
# Tag peptidici e proteici

Tag	Residues	Sequence	Size (kDa)
Poly-Arg	5-6 (usually 5)	RRRRR	0.80
Poly-His	2-10 (usually 6)	HHHHHH	0.84
FLAG	8	DYKDDDDK	1.01
Strep-tag II	8	WSHPQFEK	1.06
c-myc	11	EQKLISEEDL	1.20
S-	15	KETAAAKFERQHMDS	1.75
HAT-	19	KDHLIHNVHKEFHAAHANK	2.31
3x FLAG	22	DYKDHDGDYKDHDIDYKDDDDK	2.73
Calmodulin-binding peptide	26	KRRWKNFLAVSAANRFKKISSGAL	2.96
Cellulose-binding domains	27-189	Domains	3.00-20.00
SBP	38	MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP	4.03
Chitin-binding domain	51	TNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ	5.59
Glutathione S-transferase	211	Protein	26.00
Maltose-binding protein	396	Protein	40.00

Tag **peptidici** facilitano purificazione e rilevazione  
 Tag **proteici** aumentano livelli di espressione e solubilità,  
 facilitano purificazione e rilevazione

# Purificazione delle proteine ricombinanti con tag

Affinity tag	Matrix	Elution condition
Poly-Arg	Cation-exchange resin	NaCl linear gradient from 0 to 400 mM at alkaline pH>8.0
Poly-His	Ni <sup>2+</sup> -NTA, Co <sup>2+</sup> -CMA (Talon)	Imidazole 20–250 mM or low pH
FLAG	Anti-FLAG monoclonal antibody	pH 3.0 or 2–5 mM EDTA
Strep-tag II	Strep-Tactin (modified streptavidin)	2.5 mM desthiobiotin
c-myc	Monoclonal antibody	Low pH
S	S-fragment of RNaseA	3 M guanidine thiocyanate, 0.2 M citrate pH 2, 3 M magnesium chloride
HAT (natural histidine affinity tag)	Co <sup>2+</sup> -CMA (Talon)	150 mM imidazole or low pH
Calmodulin-binding peptide	Calmodulin	EGTA or EGTA with 1 M NaCl
Cellulose-binding domain	Cellulose	Family I: guanidine HCl or urea>4 M Family II/III: ethylene glycol
SBP	Streptavidin	2 mM Biotin
Chitin-binding domain	Chitin	Fused with intein: 30–50 mM dithiothreitol, $\beta$ -mercaptoethanol or cysteine
Glutathione S-transferase	Glutathione	5–10 mM reduced glutathione
Maltose-binding protein	Cross-linked amylose	10 mM maltose

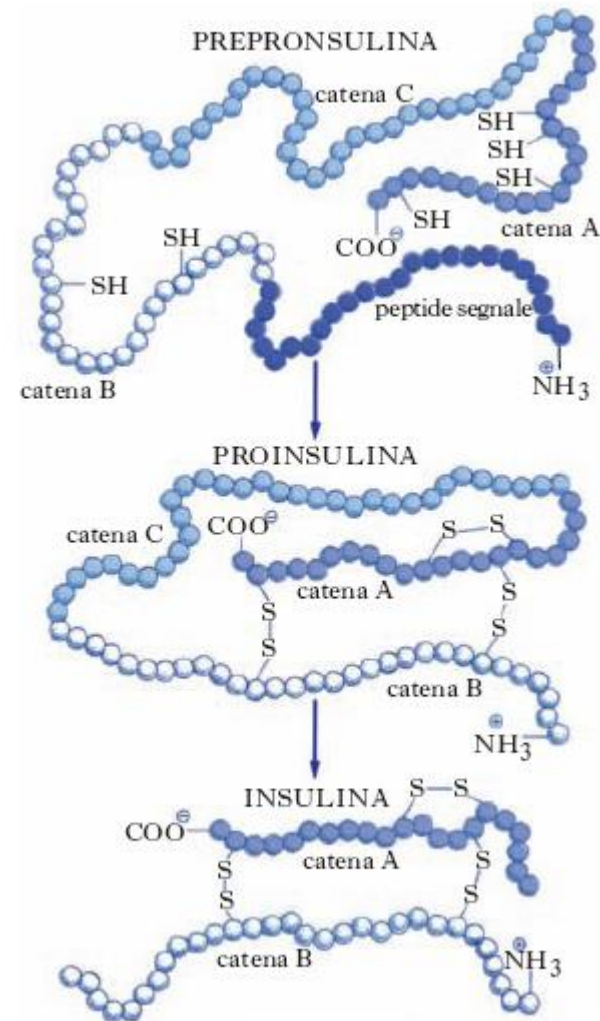




# Insulina sintetizzata nelle cellule beta del pancreas come precursore inattivo

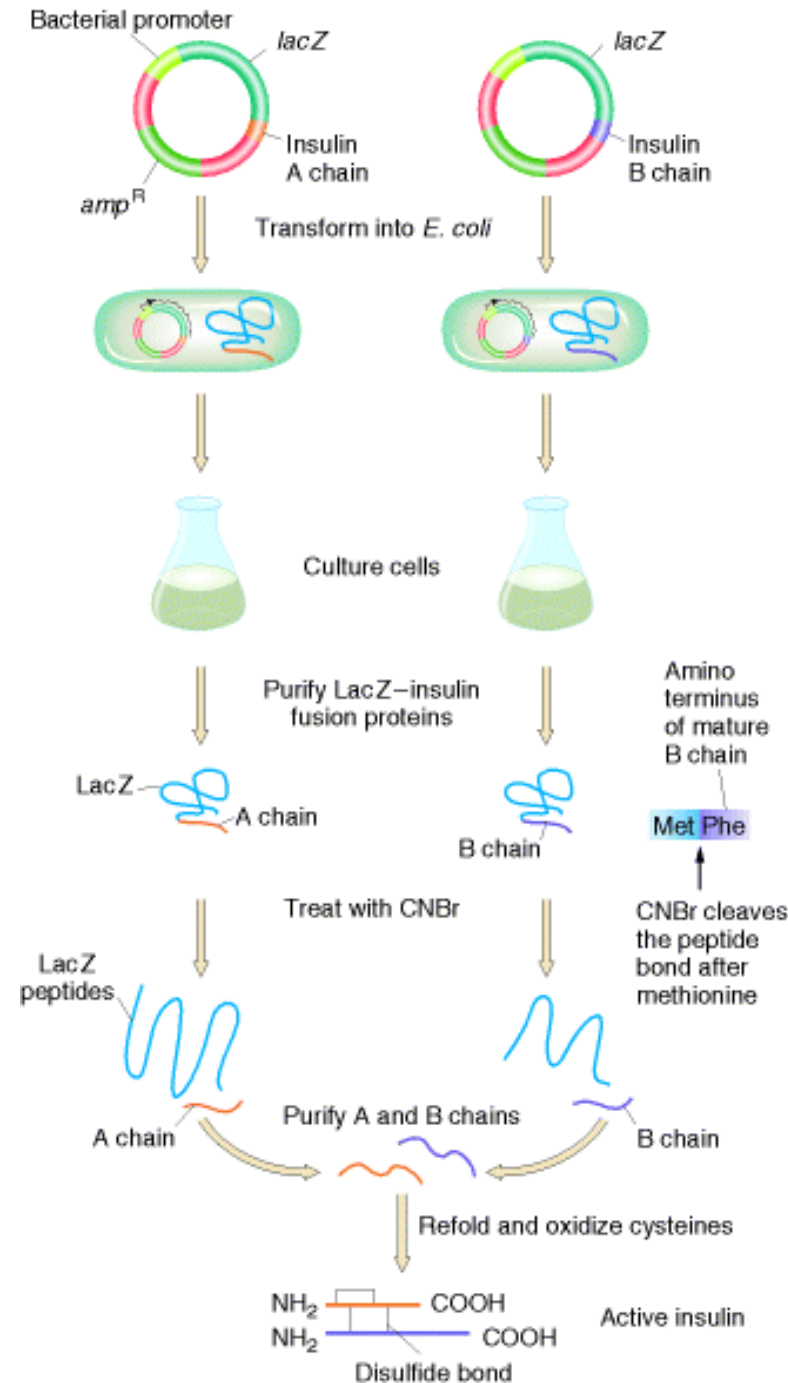
Catena A 21 aa  
Catena B 30 aa  
2 ponti S-S

Nel Golgi viene rimossa la catena C



# PRODUZIONE DI INSULINA RICOMBINANTE IN BATTERI

- Plasmidi separati codificano per Catena A e B
- Le catene sono espresse sotto il controllo di lac come proteine di fusione alla regione Nterm di LacZ
- le sequenze di LacZ sono eliminate per trattamento con bromuro di cianogeno (che taglia dopo una metionina)
- catene purificate, mescolate assieme, in opportune condizioni formano i legami S-S



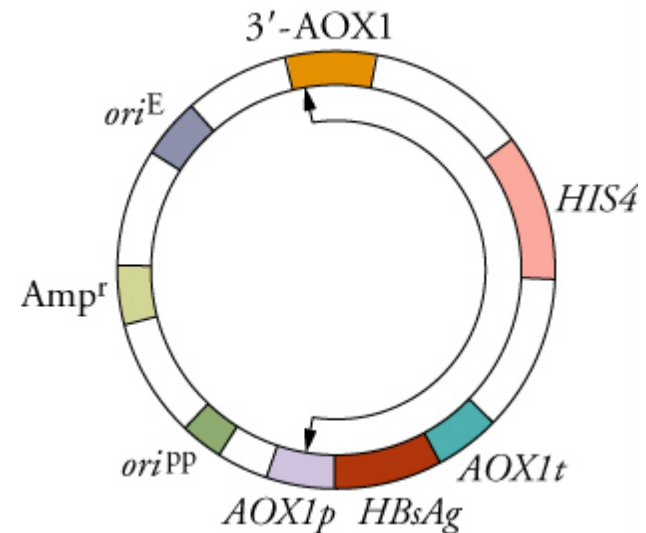


## Produzione del vaccino contro l'epatite C in lievito

- vaccini attenuati sono virus alterati in modo che non possano più riprodursi nell'organismo in cui vengono inoculati
- questi vaccini sono potenzialmente pericolosi: possono essere contaminati con virus infettivi
- i tentativi di produrre antigene di superficie del virus dell'epatite B (HBsAg ) in *E. coli* fallirono
- il gene codificante HBsAg e' stato clonato in un vettore di espressione di lievito.
- il lievito trasformato con questo vettore produce elevate quantità di HBsAg
- utilizzando fermentatori e' possibile ottenere 50-100 mg di proteina per litro di coltura

# Espressione dell'antigene di superficie del virus dell'epatite B in *Pichia pastoris* (nel citoplasma)

Promotore alcol ossidasi  
Segnali di arresto e poliadenilazione  
Regolato dalla presenza di metanolo  
Seq 3' alcol ossidasi per facilitare integrazione del DNA durante crossing over  
Gene per la biosintesi della istidina



Vettore di espressione  
che si integra in *P.*  
*pastoris*