

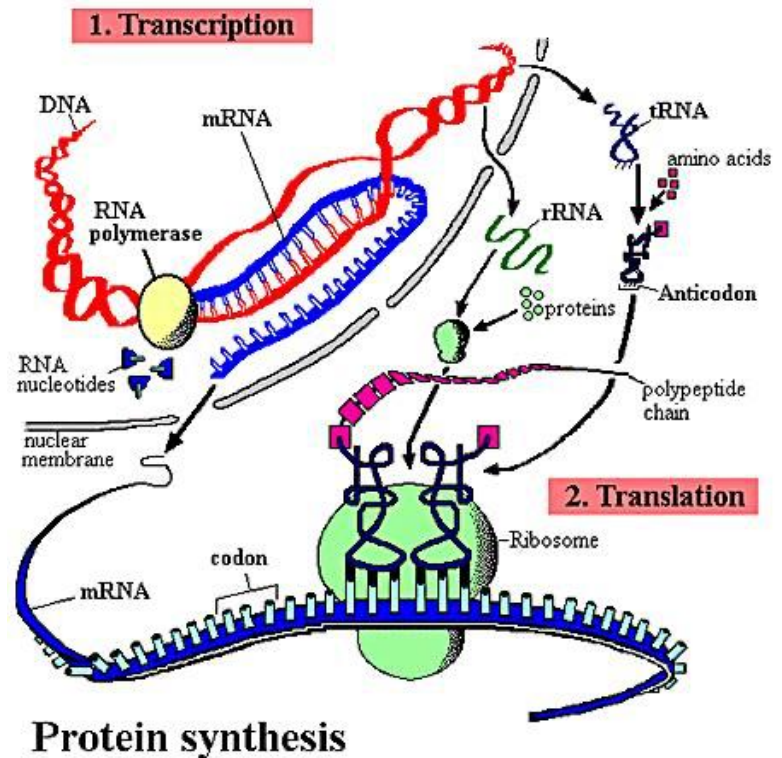
# Produzione di proteine ricombinanti

Metodi e strategie

# La biosintesi delle proteine

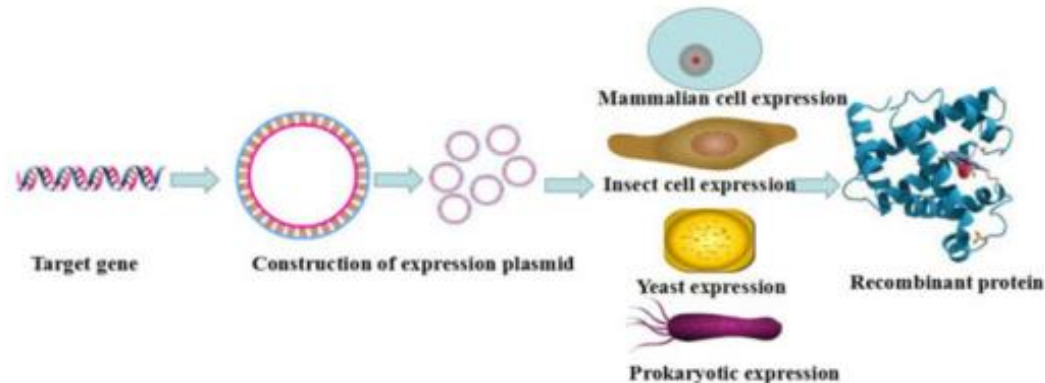
- Il DNA contiene l'informazione (sequenza)
- L'informazione è trasferita all'mRNA → **TRASCRIZIONE**
- L'mRNA è decodificato dal ribosoma per produrre la proteina → **TRADUZIONE**

Questi meccanismi e codici sono **CONSERVATI**



# Proteina ricombinante: che cosa è?

Una proteina ricombinante è una proteina che viene prodotta sfruttando la tecnologia del DNA ricombinante in un organismo eterologo (ospite) sotto il controllo di un promotore specifico per l'ospite



**Recombinant DNA:** DNA sequences from different sources are joined to produce new DNA molecules

# Proteina ricombinante: perché?

In teoria diventa possibile produrre  
quantità virtualmente **illimitate** di  
**qualsiasi** proteina

Diventa possibile **modificare** la  
sequenza codificante

# Proteina ricombinante: perché?

1. **Accademia**: studio delle relazioni struttura-funzione:
  - individuazione domini strutturali e funzionali
  - determinazione del ruolo di singoli residui amminoacidici (mutagenesi)
2. **Biofarma**: produzione di **farmaci** (es. insulina, anticorpi, ormoni, anticoagulanti, fattori della coagulazione) o antigeni per la produzione di **vaccini** (es. HBsAg)
3. **Industria**: produzione di **enzimi** e **proteine** per applicazioni industriali

Ognuna di queste applicazioni ha requisiti diversi per quanto riguarda la **qualità** e la **quantità** di proteina.

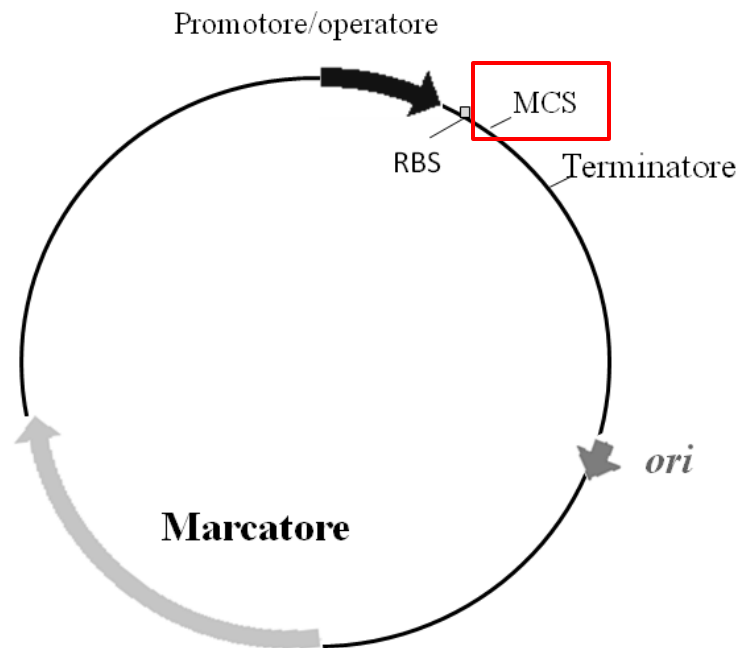
# Proteina ricombinante: come?

1. Sintesi/clonaggio sequenza codificante cDNA
  - strategie basate sulla **PCR**: è necessario avere a disposizione il DNA/RNA templato!
2. Clonaggio della sequenza codificante 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**

# Vettori di espressione

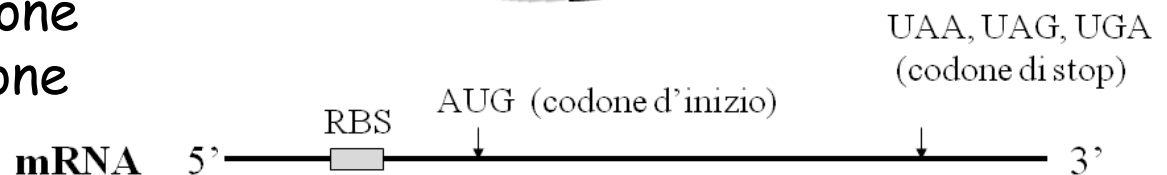
## Elementi specifici per l'ospite:

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



## Elementi per la propagazione in *E. coli*:

- Marcatore di selezione
- Origine di replicazione



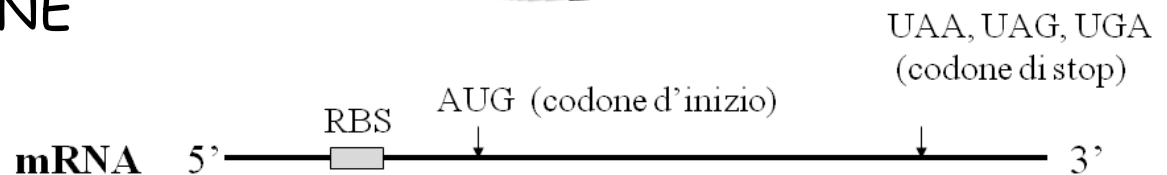
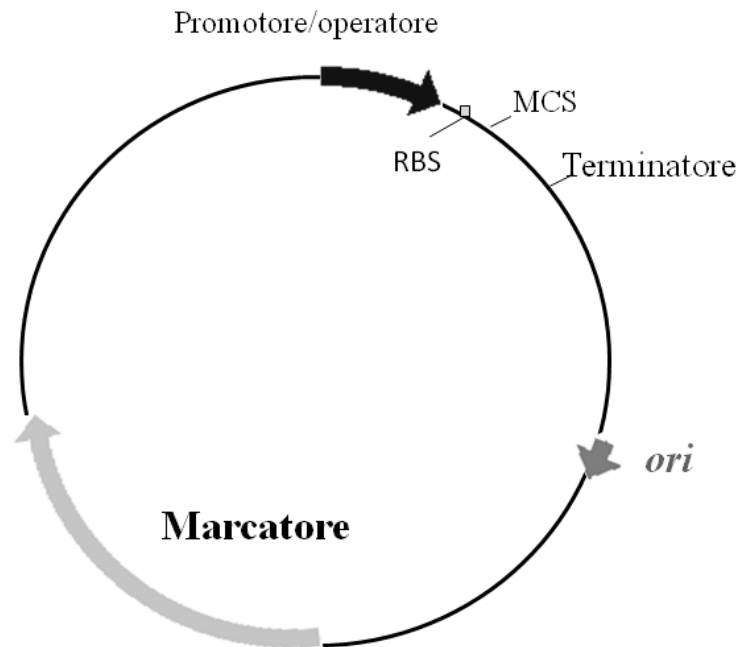
# Vettori di espressione

Elementi per modificare la sequenza codificante con aggiunta di sequenze per:

- Tag peptidici o proteici
- Segnali di localizzazione subcellulare o **secrezione**

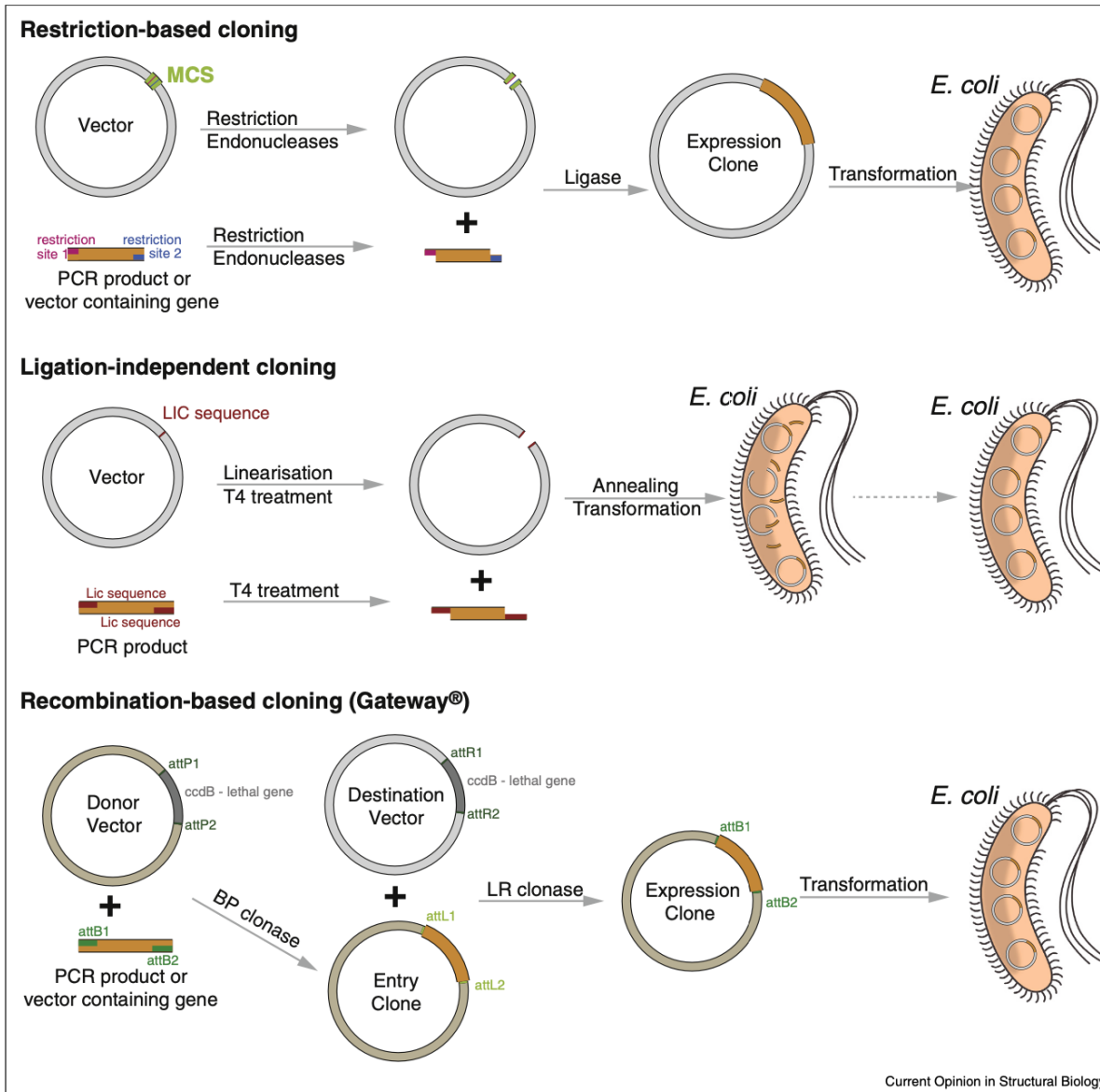


PROTEINE DI FUSIONE

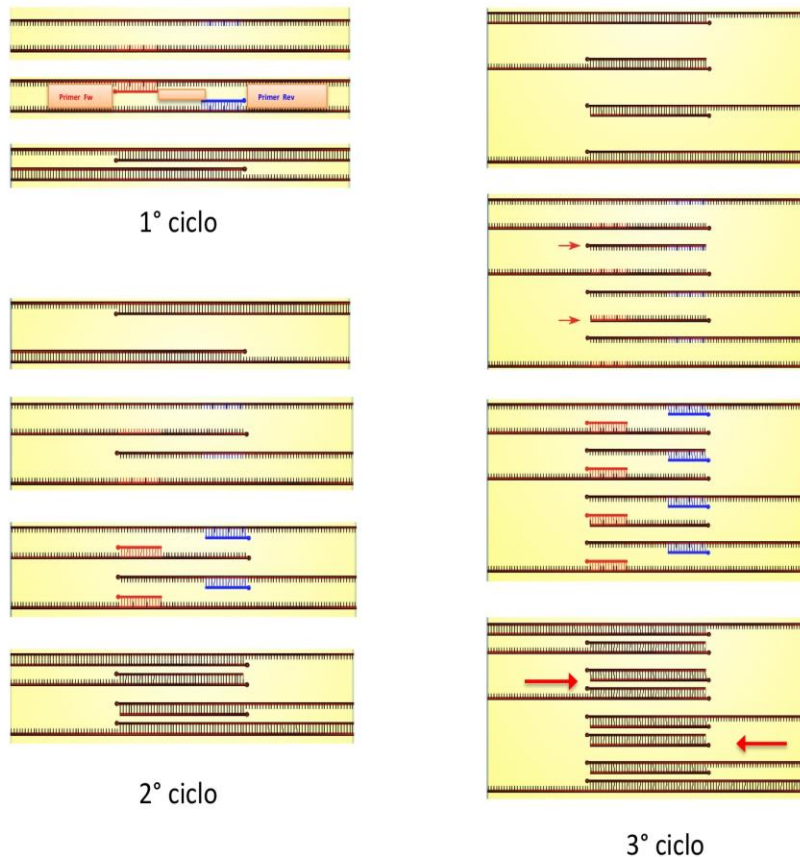




# Strategie di clonaggio di sequenze codificanti note



# Polymerase chain reaction (PCR)



Le sequenze target sono indicate dalle frecce rosse

Fasi della PCR:

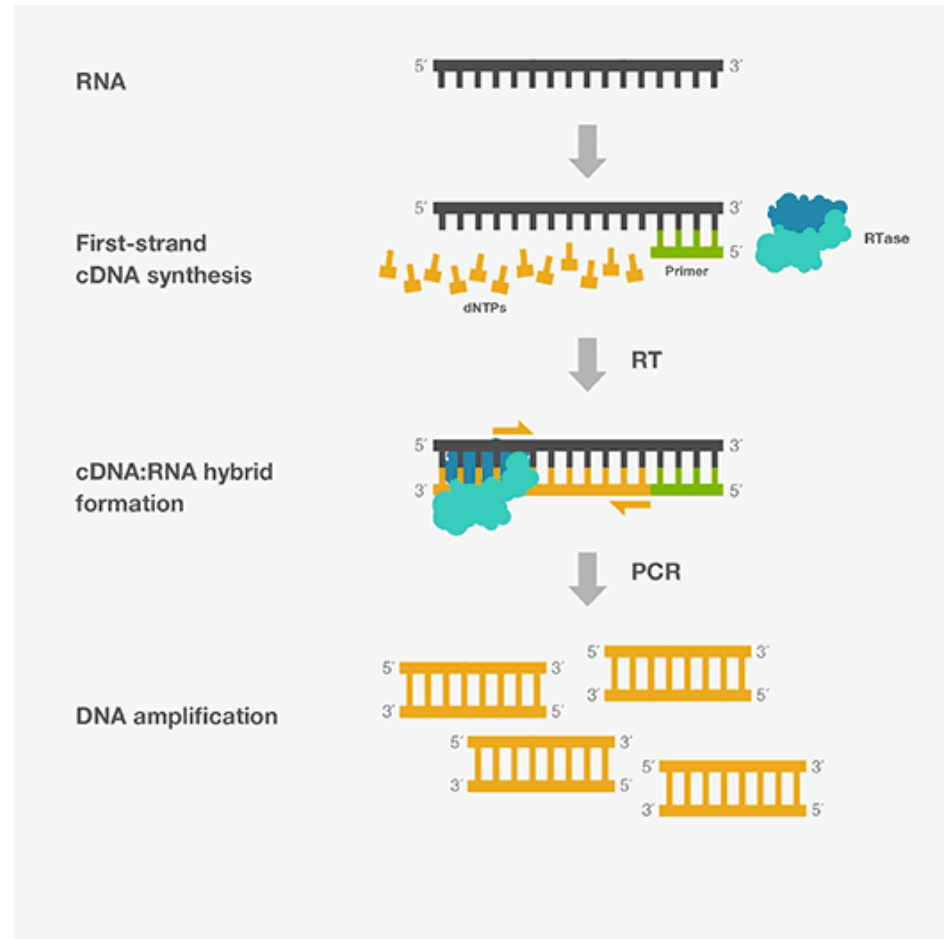
- Denaturazione
- Annealing
- Estensione

In provetta:

- DNA template
- **Due primer**
- dNTP
- Tampone +  $MgCl_2$
- Taq DNA polymerase

# Reverse transcriptase PCR (RT-PCR)

- Utilizzata per clonare sequenze codificanti eucariotiche
- I primer possono essere disegnati in modo da aggiungere sequenze utili per il clonaggio nel vettore
- Stampo: mRNA
- Enzimi: trascrittasi inversa (RT) e Taq





# Strategia di clonaggio mediante PCR se la sequenza codificante è nota

## 2. Progettazione degli oligonucleotidi (primer)

Sito di clonaggio multiplo del vettore

*BamHI* *EcoRI* *SacI* *SalI* *HindIII* *NotI* *XhoI*  
ATCGGAATTAATTC**GGATCCGAA**TTCGAGCTCC**GTCGACAAGCTT**GCGGCCGC**ACTCGAG**CACCACCACCACCACCA

Sequenza degli oligonucleotidi per il clonaggio del gene per PCR

BamHI Forward 5' -CAGGGATCC**ATGA**ACAACCTTTAATCTGCAC-3'  
HindIII Reverse 5' -CAGAAGCTT**TTA**GCGGGCGGCTTCGTATATAC-3'

Gli oligonucleotidi contengono delle **estensioni al 5'** che introducono siti di restrizione utili per il clonaggio nel vettore

# Strategia di clonaggio mediante PCR se la sequenza codificante è nota

Sequenza degli oligonucleotidi per il clonaggio del gene per PCR

BamHI Forward 5'-CAGGGATCCATGAACAACCTTTAATCTGCAC-3'  
HindIII Reverse 5'-CAGAAGCTTTTAGCGGGCGGCTTCGTATATAC-3'



PCR su DNA genomico oppure RT-PCR su RNA

Digestione del prodotto di PCR con gli enzimi di restrizione  
BamHI e HindIII

Ligazione nel vettore digerito con BamHI e HindIII

Trasformazione di *E. coli*, estrazione del DNA plasmidico e conferma  
della presenza della sequenza del gene di interesse

# Strategie di clonaggio se la sequenza codificante NON è nota

## Screening di librerie di cDNA o di DNA genomico

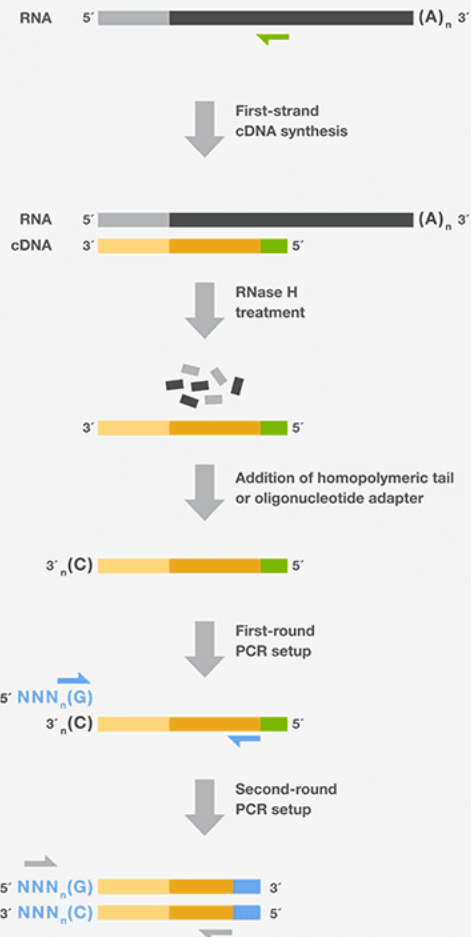
1. ibridazione della libreria con una sonda a DNA gene-specifica
2. identificazione e recupero del plasmide/fago che contiene la sequenza di interesse

## RACE (Rapid Amplification of cDNA Ends)

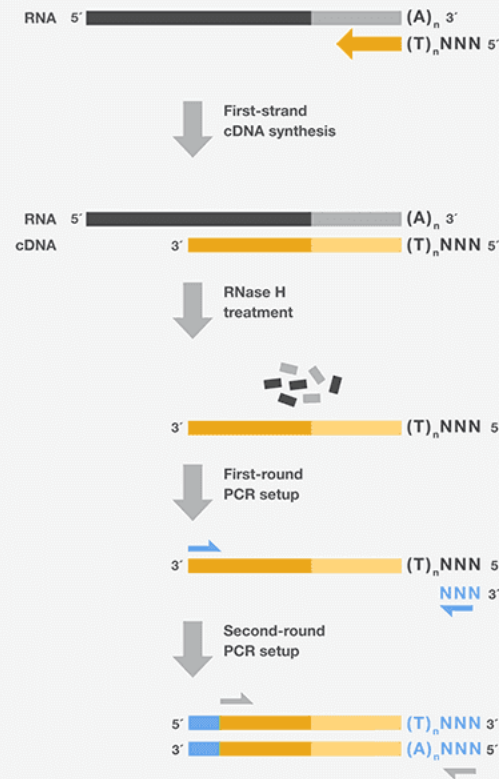
5' RACE

3' RACE

## A 5' RACE



## B 3' RACE



Se il primer gene-specifico  
appaia sul codone d'inizio  
ATG si ottiene l'intera  
sequenza codificante



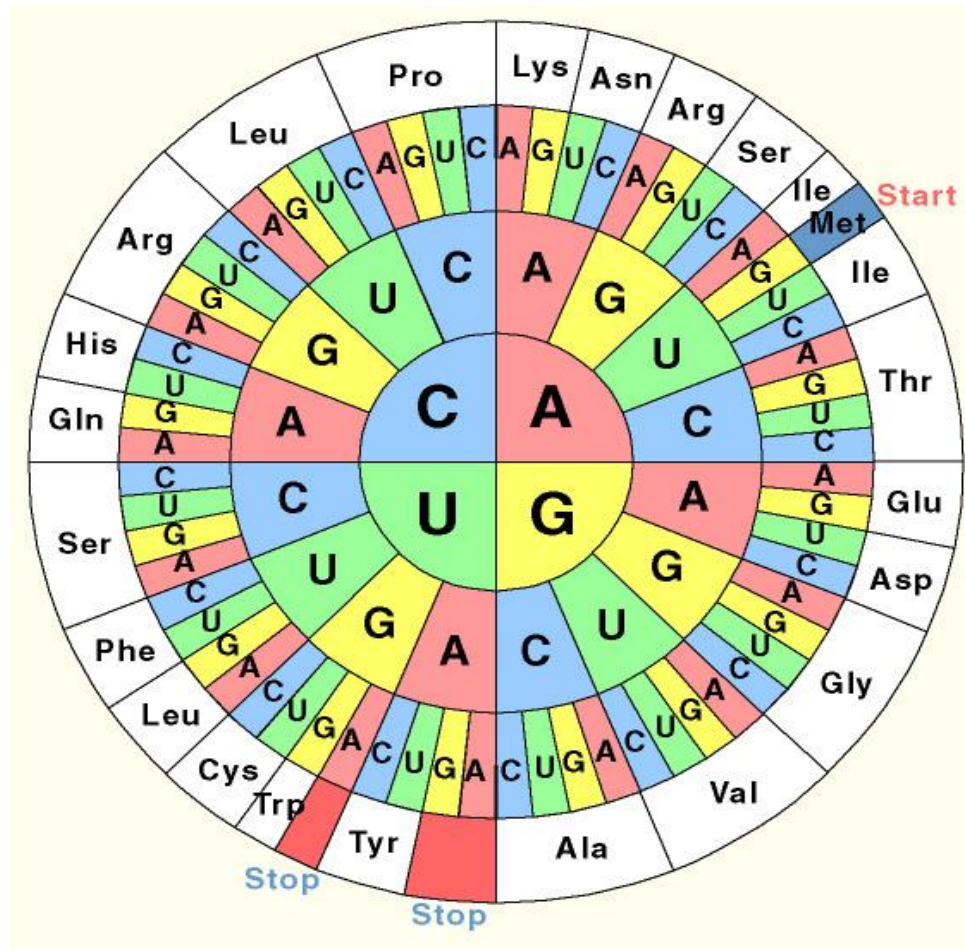
E' necessario avere informazioni sulla sequenza aminoacidica e/o nucleotidica per progettare gli oligo gene-specifici

Le informazioni si possono ottenere:

- direttamente sulla proteina (=> degradazione di Edman o spettrometria di massa)
- mediante analisi bioinformatiche (=> allineamenti di sequenze ed identificazione di sequenze conservate)

# COME PROGETTARE UN OLIGONUCLEOTIDE A PARTIRE DA UNA SEQUENZA AMINOACIDICA

Il problema della degenerazione del codice genetico:



# COME PROGETTARE UN OLIGONUCLEOTIDE A PARTIRE DA UNA SEQUENZA AMINOACIDICA

La degenerazione del codice genetico riduce la probabilità di un appaiamento preciso dell'oligo!

Evitare aminoacidi codificati da sei codoni (Leu, Ser, Arg)  
Escludere la terza base dell'ultimo codone dal conteggio  
Scegliere la sequenza aminoacidica a minore degenerazione

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

La degenerazione può essere ulteriormente ridotta

1) con l'introduzione di basi insolite

deossiinosina (I)    N→I

I→A    I→G

I→C

I→T

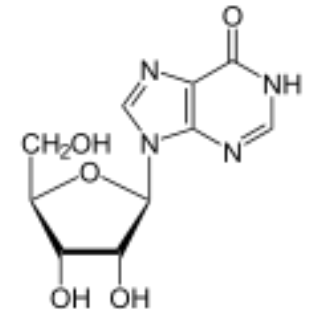
ok    no

es. Arg

CGN

AGA/G

5' -IGN-3'



2) Con l'utilizzo del 'codon usage'

- non tutti i codoni per un dato aminoacido sono utilizzati con la stessa frequenza. Scegliendo il codone (o i due codoni) più usato(i) per ciascun residuo aminoacidico nella sequenza bersaglio si può ridurre la degenerazione totale dell'oligonucleotide, ma nello stesso tempo si riduce la probabilità di un appaiamento perfetto.
- si possono disegnare oligonucleotidi più lunghi

# Frequenze di uso dei diversi codoni 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

# Un esempio: la progettazione dell'oligo per il clonaggio della proteina Ftr1 del lievito *Pichia pastoris*

Allineamento delle sequenze aminoacidiche di Ftr1 di vari lieviti per identificare regioni conservate

```
FTR1_YEAST      MPNKVFNVAVFFVVFRECLEAVI VISVLLSFLKQA IGEHDR-----ALYRKLRIQVWVGW 55
FTR2_CANAL     -MVDVFNQVFFVVFREALEAVI VVSVLLAFVKQSMGSSANG---PELKKKLYRQIWLGA 56
FTR1_CANAL     -MVDVFNQVQIFFIVFRESLEAI I VVSVLLAFVKQSMGGSSD---PQLKKRLYRQIWLGA 55
YAK7_SCHPO     MAKDVFSVAIFFIVLRETLEASI IVSVLMSFISQTLMDKDGNVTDPKLKRKFMQLQVWIGS 60
      .**.* :**.*:** ** * :***:*.**.:
FTR1_YEAST      LLGFIIICLAIGAGFI GAYYSLQKDI FG-SAE DLWEGI FCMIAT IMISMGI PMLRMNKM 114
FTR2_CANAL     GLGVLICIIIGGAFI GTFYGLGKDIWG-KSE DLWEGI FCI IATVLITAMGIPMLRINKMK 115
FTR1_CANAL     GLGVLVCLY-GVLSIGASYGLGKDI FGVISEDLWEGI FCI IATVLITAMGIPMLRINKMK 114
YAK7_SCHPO     FTALFICLAIGGGFI GAFYALDKDIWS-GSEEI WEGV FSLI AVLITVMGFAMLRVSHLQ 119
      ..:.*: * **:*.* ***: . :*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
FTR1_YEAST      SKWRVKIARSLVEI PHRKRDY FKIGFLSRRYAMFLLPFI TVLREGLEAVV FVAGAGIT 174
FTR2_CANAL     EKWRVKLAQALIKSPENKKNR FKLGYLGKKYALFILPFI TCLREGLEAVV FVGGVGITSP 175
FTR1_CANAL     EKWRVKLAQALIKSP TNKKDR FKLGYLGKKYALFILPFLQVLREGLEAVV FVGGVGLNSP 174
YAK7_SCHPO     EKWRKKLMKSIAN--RKAAG---ISNWGKKYSMFLLPFFTVLREGLEVVV FVGGVGLETP 174
      .*** * : : : : . : . : . : . : * : * : * : * : * : * : * : * : * : * :
FTR1_YEAST      GSHASAYPLPVVGLICGGLVGYLLYYGASKSSLQIFLI LSTSI LY LISAGLFSRGAWYF 234
FTR2_CANAL     ---ASSFPPIPVI VGIICGLAVGALLYFGSNMSMQIFLI ISTCILYLI AAGLFSRGVWFF 232
FTR1_CANAL     ---ATSPPIPVI VGLIAGIVVGALLYFGS SMSMQIFLI ISTCILYLI AAGLFSRGIWYF 231
YAK7_SCHPO     ---ATAFPLPVI CGLIVGCLIGFY IYRGGVMNLQWFLI ASTCILYLI SAGLMSKATFYF 231
      * : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :
FTR1_YEAST      ENYRFNLAGGDASEGGDGNGSYN IRKAVYHVNCCNPELDN--GWDIFNALLGWQNTGYL 292
FTR2_CANAL     ESYQYNKKTGGDAAENGS GPGTYD ISKSVWHVNCCNPLTDN--GWDIFNAILGWQNSATY 290
FTR1_CANAL     ETNTYNKKTGGDASENGS GPGTYD ISKSVWHVNCRN PETDN--GWDIFNAILGWQNSATY 289
YAK7_SCHPO     EMNKWNHQTGGDAGELGDGPGSYP FKS AVWHVNYGN PEMNSNGGYMIFNAILGWNNTGTY 291
      * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :
FTR1_YEAST      SSMLCYNIYWLVI IIVLSLMI FEERRGHL PFTKNLQLKHLNPGYW IKNKKQELTEEQKR 352
FTR2_CANAL     GSVISYNVYWIFI IISVLLLMVYEEKHGHLPLTKNLT LVQLNPMYH IKGKKLLELNKAEKE 350
FTR1_CANAL     GSVISYNIYWLFI ICVLLLMVYEEKHGHL PFTKNLT LVQLNPMYH IKGKKLLELNKAEKD 349
YAK7_SCHPO     GSI LSYIIYWL FVAF IMFLMWYKERRAARLLIAKLGDKVVDLEAASSHTPVQSSSEDEF 351
      .*:.*. * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :
FTR1_YEAST      QLF AKMENI NFNEDGE INVQENYELPEQT TSHS SSQNVATDKEVLHVKADSL 404
FTR2_CANAL     ELFTKLQKQKFGDVQE IDETS SNKLVE TQENK----- 382
FTR1_CANAL     ELFTKLQQQNFQAAEVD ETS SNKWMSQENS----- 381
YAK7_SCHPO     KINSPTDDKGDKAIDIVTEVRESSSPVEEHKDDKTVDVINEIRESH----- 397
      : : : . : . . . . . . . . . . . . . . . . . . . . . . . . . . . .
```

# Progettazione dell'oligonucleotide per il clonaggio del cDNA della proteina Ftr1 del lievito *Pichia pastoris*

Sequenza scelta: Leu-Arg-Glu-Gly-Leu-Glu-Ala-Val-Val

- conservata in tutte le proteine Ftr1
- i due residui di Glu sono essenziali per la funzione di Ftr1

Sequenza scelta: Leu-Arg-Glu-Gly-Leu-Glu-Ala-Val-Val

Degenerazione: 6 6 2 4 6 2 4 4 4

Codon usage	Leu	TTA 0.16	TTG 0.33				
		CTT 0.17	CTC 0.08	CTA 0.12	CTG 0.16		
	Arg	AGA 0.47	AGG 0.16				
		CGT 0.16	CGC 0.05	CGA 0.11	CGG 0.05		
	Glu	GAA 0.58	GAG 0.42				
	Gly	GGT 0.43	GGC 0.14	GGA 0.32	GGG 0.10		
	Ala	GCT 0.45	GCC 0.25	GCA 0.24	GCG 0.06		
	Val	GTT 0.42	GTC 0.23	GTA 0.16	GTG 0.20		

Oligo: TT(AG) AGA GA(AG) GG(AT) (CT)T(AG) GA(AG) GC(ACT)  
GT(AGCT) GT

Degenerazione 768 invece di 221184

# Clonaggio del cDNA della proteina Ftr1 del lievito *Pichia pastoris*

## 3' RACE:

1. RT con oligo poli-dT-linker su mRNA di *Pichia pastoris*
2. PCR con oligo Ftr1 forward e linker (*GACTCGAGTCGACATCG*)

Risultato ottenuto: prodotto di PCR di circa 900 bp

TTAAGAGAGGGTTTAGAGGCTGTAGTGTTCATAGGTGGTGTGGCCTGGGATCTCCAGCTACATCTTTTCCCATCCCTG  
TGATTGTTGGTCTTATTGCTGGTATTTCTGTGCGGTGTGCTTTTGTACTATTCAGGATCAACTCTTTCTTTGCAAGTTTT  
CTTATGTATATCGACGGCTATTTTGTACTTGATTGCTGCTGGGTTGTTTTCTAGGGCTATCTGGTTCTTTGAAACCTAC  
AAATACAACCAAAAGACTGGAGGAGATGCTTCTGAAAATGGTTCGGGACCCGGTACTTATGACATTAAGACTTCTGTTT  
GGCATGTGAACTGCTGTAACCCTGAAACTGATAATGGTGGGATATTTTTAACGCGTTGTTGGGATGGCAAAACTCTGC  
CACTTATGGCTCAGTCATAGGCTACAACATATACTGGCTGGCGGTAATGATCACTTTGTATTTACTCTGGTTCGAAGAA  
AAGAATAACCACTTGCCATTTATGAAAAATCTGAAGTTGAGACAACCTGAACCCATTATACTGGATGAAGGGTAAAAACA  
AGAAGGAGGTCTCCAAGGAAGATCAGGAAAAATTATTTGAGCAATTGAAGAGTAAGGAGTTGCCAATAAACTTGCGGA  
GGAAT**TAAT**TTAGATTGACCTAGCTGAATCTTGTTAGCAACTGTATGCACACACACACACACACACACACACATAACAAG  
GGCTAACAGAGGATGATACGTTTTTTCAGTCACAAACACATACACATGCACATGTACCCTAACAAAGAACAAGACATTTTG  
CTTGAATTAGAATCAGGACTGCAAAAGGTTTACCGCATACTACGTACGCACATAACCCATCTAATACACACGTTTTGT  
ATAAATCGATGTCGACTCGAGTC

Oligo:            **TT(AG) AGA GA(AG) GG(AT) (CT)T(AG) GA(AG) GC(ACT) GT(AGCT) GT**  
Prodotto PCR: TTA    AGA GAG    GGT    TTA            GAG    GCT    GTA            GT



# Allineamento della sequenza aminoacidica dedotta di Ftr1 di *Pichia pastoris* con le altre permeasi Ftr1

```

PpaFTR1      LREGLEAVVFVGGVGLGSP---ATSFPIPVIVGLIAGISVGVLLYYSGSTLSLQVFLCI
CaFTR2      LREGLEAVVFVGGVGITSP---ASSFPIPVIVGIICGLAVGALLYFSGNMSMQIFLII 212
CaFTR1      LREGLEAVVFVGGVGLNSP---ATSFPIPVIVGLIAGIVVGALLYFSGSSMSMQIFLII 211
FTR1_YEAST  LREGLEAVVFVAGAGITTOGSHASAYPLPVVVGGLICGGLVGYLLYYGASKSSLQIFLIL 214
YAK7_SCHPO  LREGLEVVFVGGVGLETP---ATAFPLPVICGLIVGCLIGYFIYRGGNVMNLQWFLIA 211
*****.***.* * : *::*:::: *:* * . * ::* .. ::* **

PpaFTR1      STAILYLIAAGLFSRAIWFFETYKYNQKTGGDASENGSGPGTYDIKTSVWHVNCNPET
CaFTR2      STCILYLIAAGLFSRGVWFFESYQYNKKTGGDAAENGSGPGTYDISKSVWHVNCNPLT 271
CaFTR1      STCILYLIAAGLFSRGIWYFETNTYNKKTGGDASENGSGPGTYDISKSVWHVNCRNPE 270
FTR1_YEAST  STSILYLISAGLFSRGAWYFENYRFNLSGGDASEGGDNGSYNIRKAVYHVNCNPEL 273
YAK7_SCHPO  STCILYLISAGLMSKATFYFEMNKWNHQTGGDAGELGDGPGSYPFKSAVWHVNYGNPE 270
**.*****.***.*** ::* * .****.* *.* *:* : :*:* **

PpaFTR1      DN--GWDIFNALLGWQNSATYGSVIGYNIYWLAVMITLYLLWFEEKNNHLPFMKNLKL 329
CaFTR2      DN--GWDIFNAILGWQNSATYGSVISYNYWIFII SVLLLMVYEEKHGHLPFTKNLTL 328
CaFTR1      DN--GWDIFNAILGWQNSATYGSVISYNIYWLFIICVLLLMVYEEKHGHLPFTKNLTL 328
FTR1_YEAST  DN--GWDIFNALLGWQNTGYLSSMLCYNIYWLVLII VLSLMI FEERRGHLPFTKNL 331
YAK7_SCHPO  NSNGGYMIFNAILGWNNTGTYGSILSYIIYWLVFVAFIMFLMWYKERRAARLLIAKL 330
:. * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

PpaFTR1      LNPLYWMMKGKKNKKEVSKEDQEKLFEQLKSKEF-----ANKLAEE-----
CaFTR2      LNPYHIKGGKKLELNKAKEEELFTKLQKQKFGDVQEIDETSSNKLIVETQENK----- 382
CaFTR1      LNPYHIKGGKKLELNKAKEDELFTKLQQNFGQAAEVDETSSNKWMDSQENS----- 381
FTR1_YEAST  LNPGYWIKNKKQELTEEQKRQLFAKMNINFNEDGEINVQENYELPEQTTSHSSQNVA 391
YAK7_SCHPO  VDLEAASSHTPVQSSSEDEFKINSPTDDKGDKAIDIVTEVRESSSPVEEHKDDKTVDVI 390
. . . . .
FTR1_YEAST  TDKEVLHVKADSL 404
YAK7_SCHPO  NEIRESH----- 397

```

Per completare la sequenza codificante di Ftr1 di *Pichia pastoris* il frammento ottenuto è stato utilizzato come sonda nello screening di una libreria di cDNA

# Modificazione della sequenza codificante.

## Strategie di mutagenesi basate sulla PCR

### Mutagenesi sito-specifica

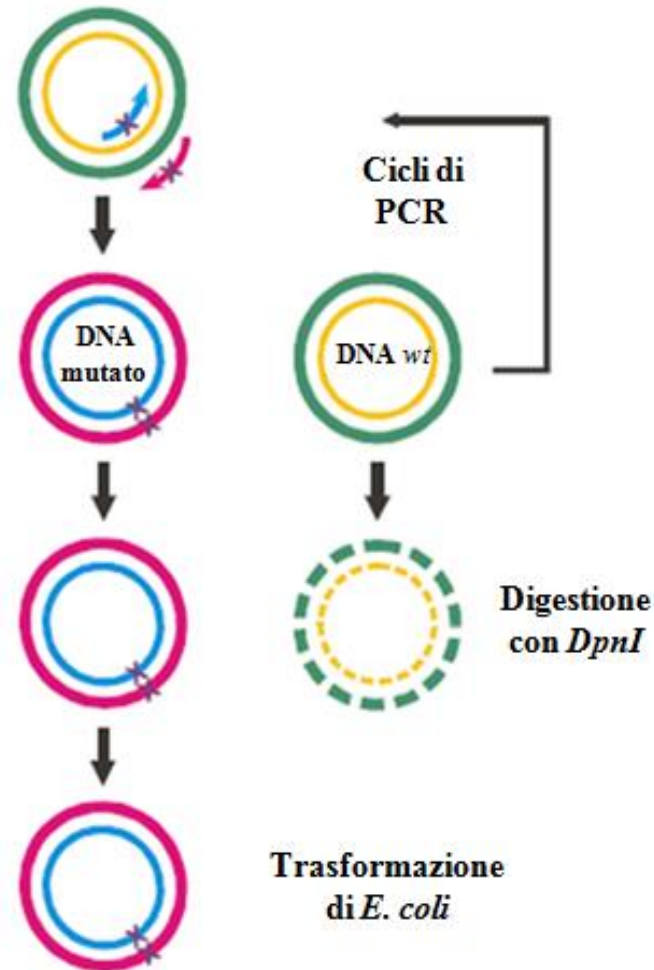
per valutare il ruolo strutturale/funzionale di specifici aminoacidi, si producono poche varianti sulla base di ipotesi predefinite

### 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 → è necessario un metodo di screening efficiente

## Mutagenesi sito-specifica **QuikChange**:

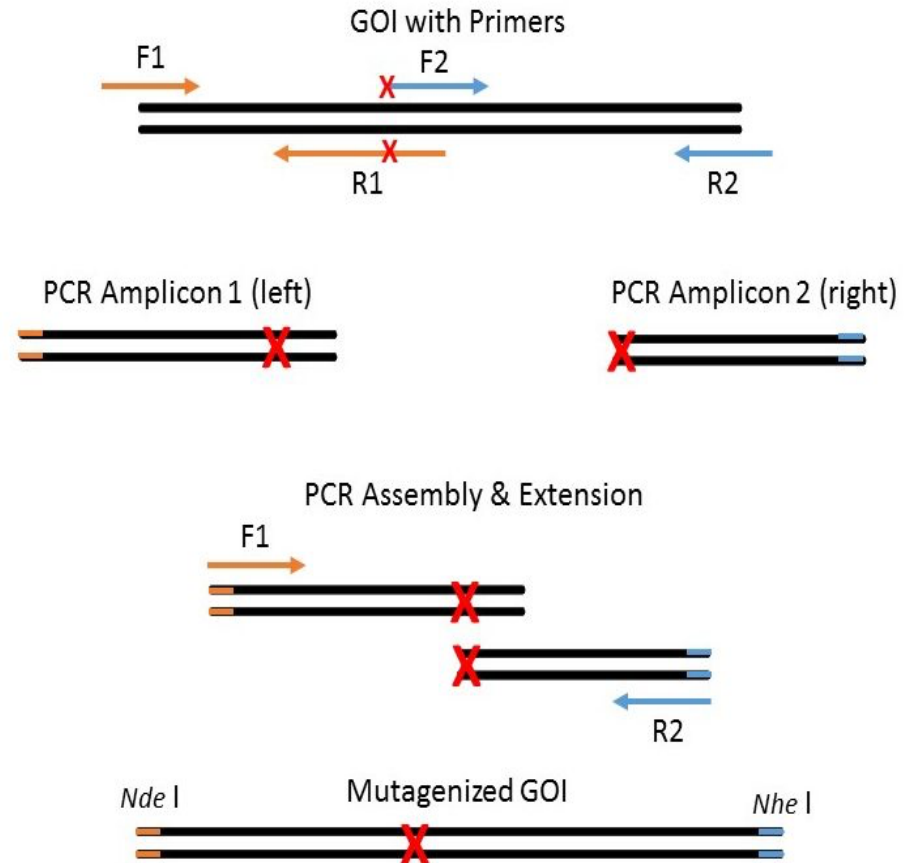
1. PCR di amplificazione dell'intero plasmide con due oligo complementari che portano la mutazione
2. Digestione con DpnI per degradare il template wild type (metilato)
3. Trasformazione di *E. coli*
4. Verifica mediante sequenziamento del DNA



# Mutagenesi sito-specifica: overlap extension PCR

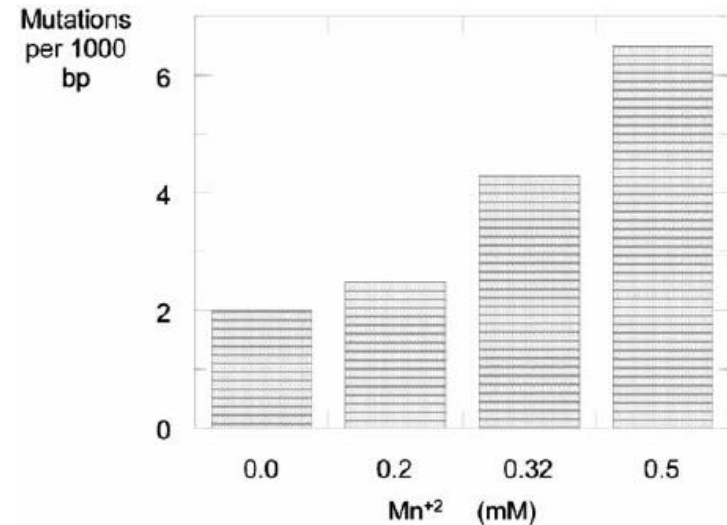
Gli oligo mutagenici vengono utilizzati insieme a oligo alle estremità per amplificare porzioni di sequenza codificante che vengono poi usate come template in una seconda PCR che riassume il cDNA

## Design of the Mutagenesis by PCR Overlap Extension

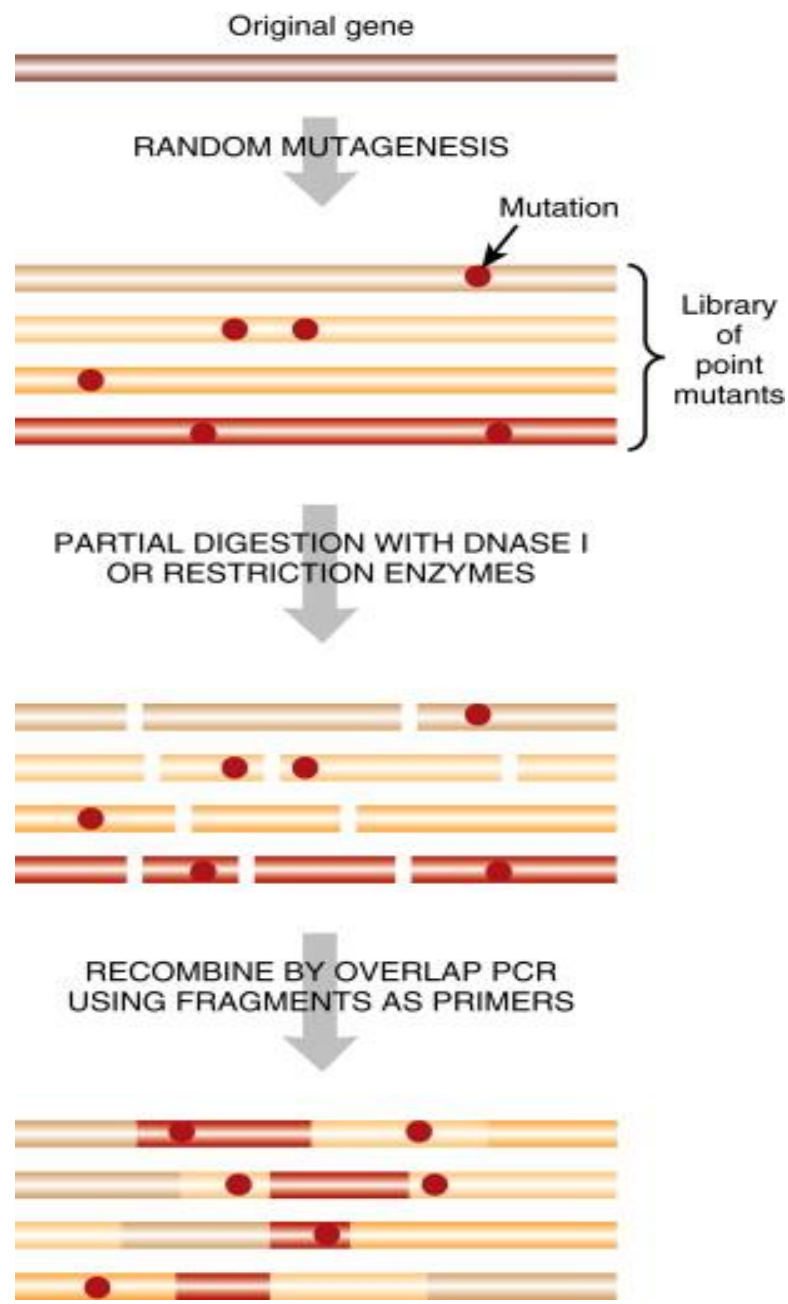


# Evoluzione guidata in vitro: error-prone PCR

- 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 suo posto un dNTP diverso),
  - 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)

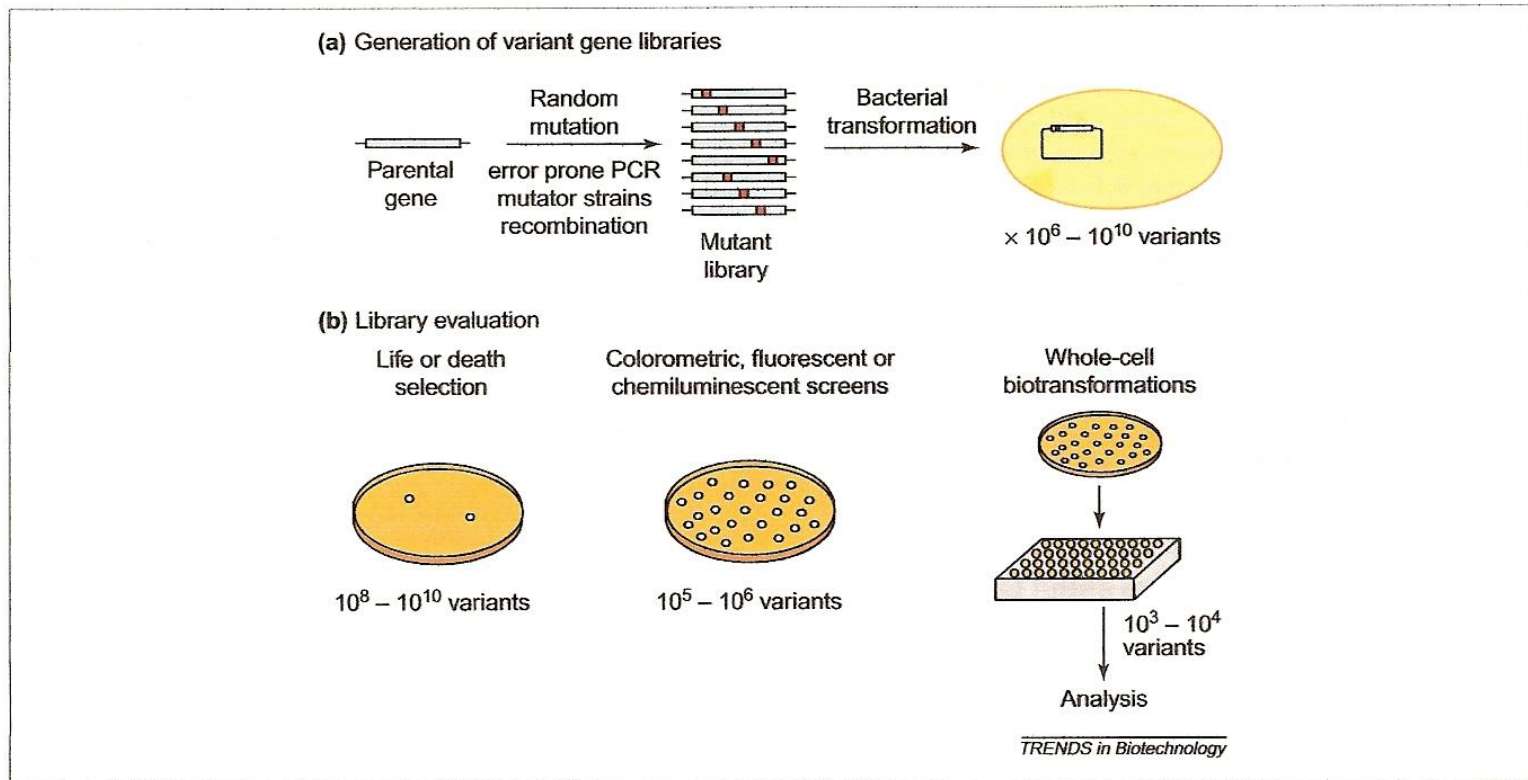


Evoluzione guidata in vitro:  
**DNA shuffling**



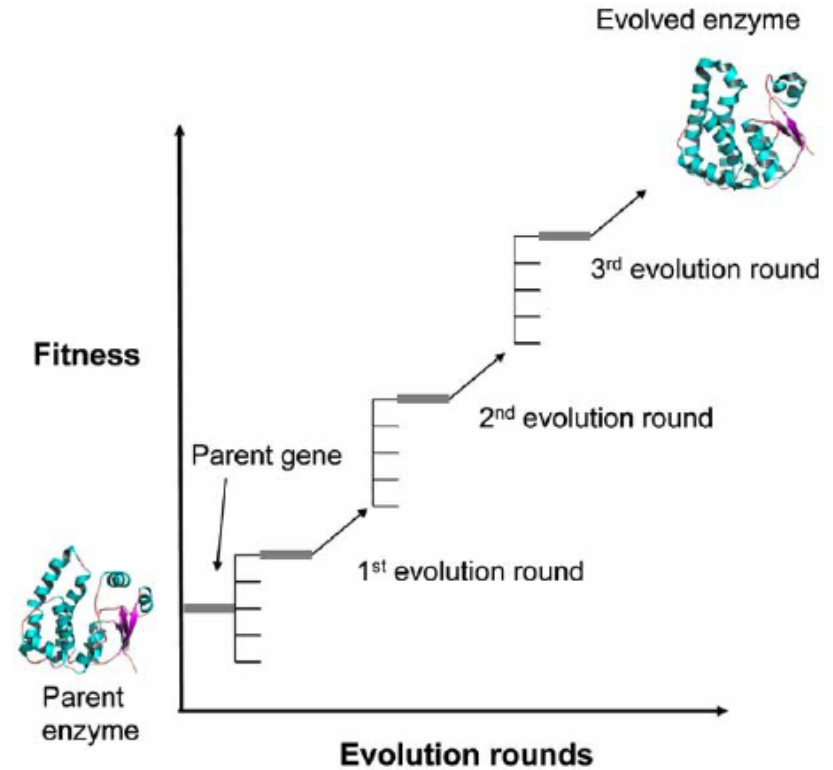
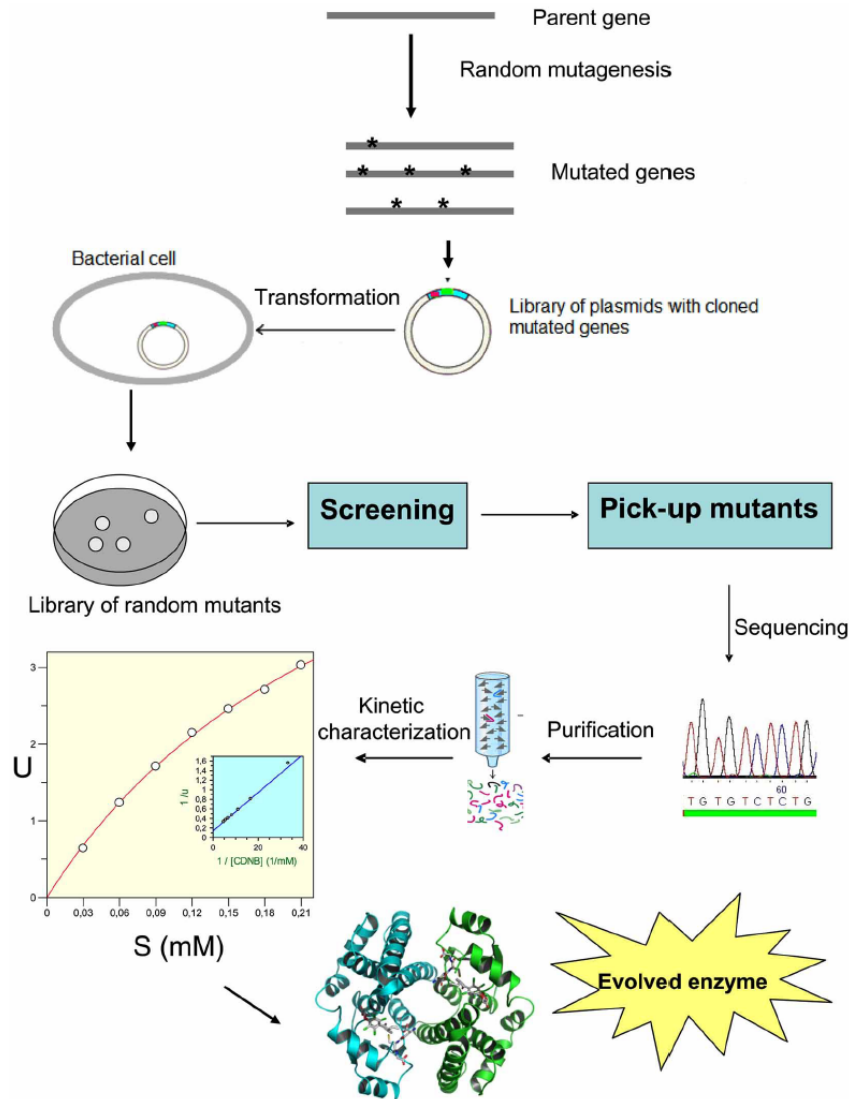
# Evoluzione guidata in vitro

## Metodi di screening delle librerie di varianti



**Figure 1.** Strategies for the directed evolution of enzymes involving (a) generation of libraries of variant genes and (b) high-throughput screening of the libraries using different analytical methods.

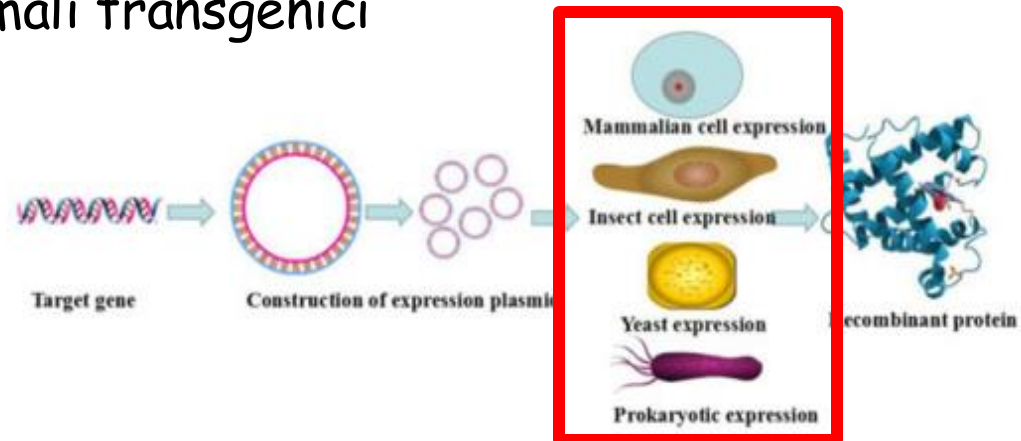
# Evoluzione guidata in vitro Strategia





# Organismo ospite

- **Procarioti**
  - Batteri (*Escherichia coli*, *Bacillus* sp.)
- **Eucarioti**
  - Lievito (*Saccharomyces cerevisiae*, *Pichia pastoris*, *Kluyveromyces lactis*)
  - Cellule di mammifero (CHO, BHK, HEK-293)
  - Cellule di insetto (Sf9)
  - Piante e animali transgenici



# Scelta del sistema di espressione

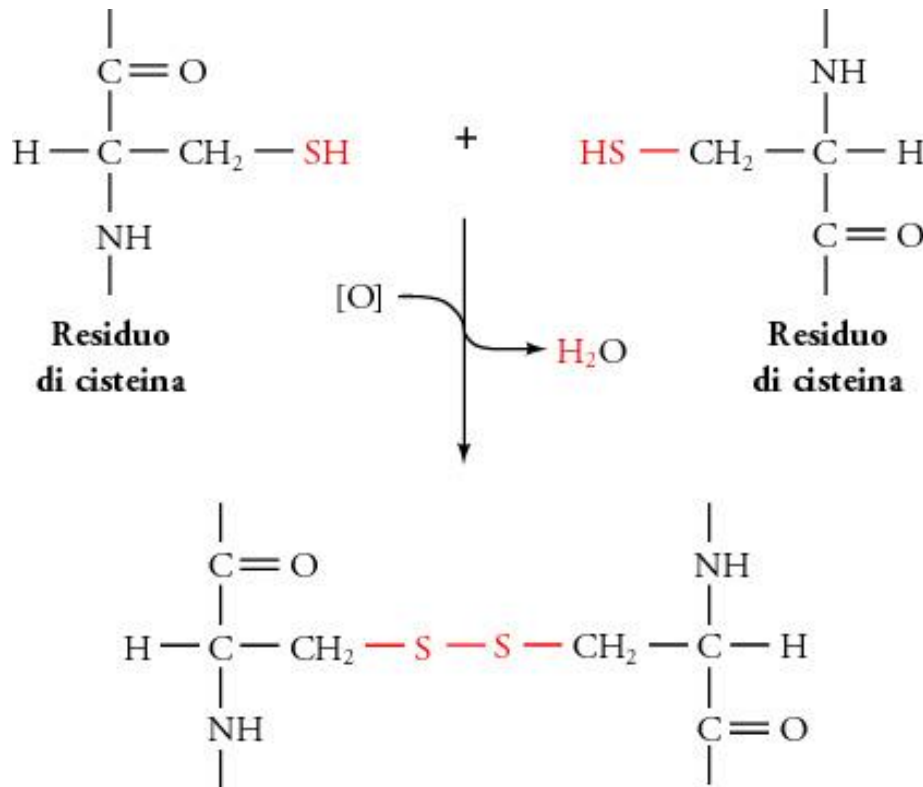
- Caratteristiche della proteina, incluse eventuali modifiche post-sintetiche necessarie per il folding e l'attività biologica
- Tempo di produzione
- Costo di produzione
- Regolamentazione per l'approvazione
- Diritti

**Sistema di espressione:** combinazione ospite/vettore (e modalità) di espressione

# Principali modifiche post-sintetiche delle proteine

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

# Cisteina e ponti disolfuro



La formazione di ponti disolfuro avviene nel **periplasma** nei batteri e nel **reticolo endoplasmatico** negli eucarioti

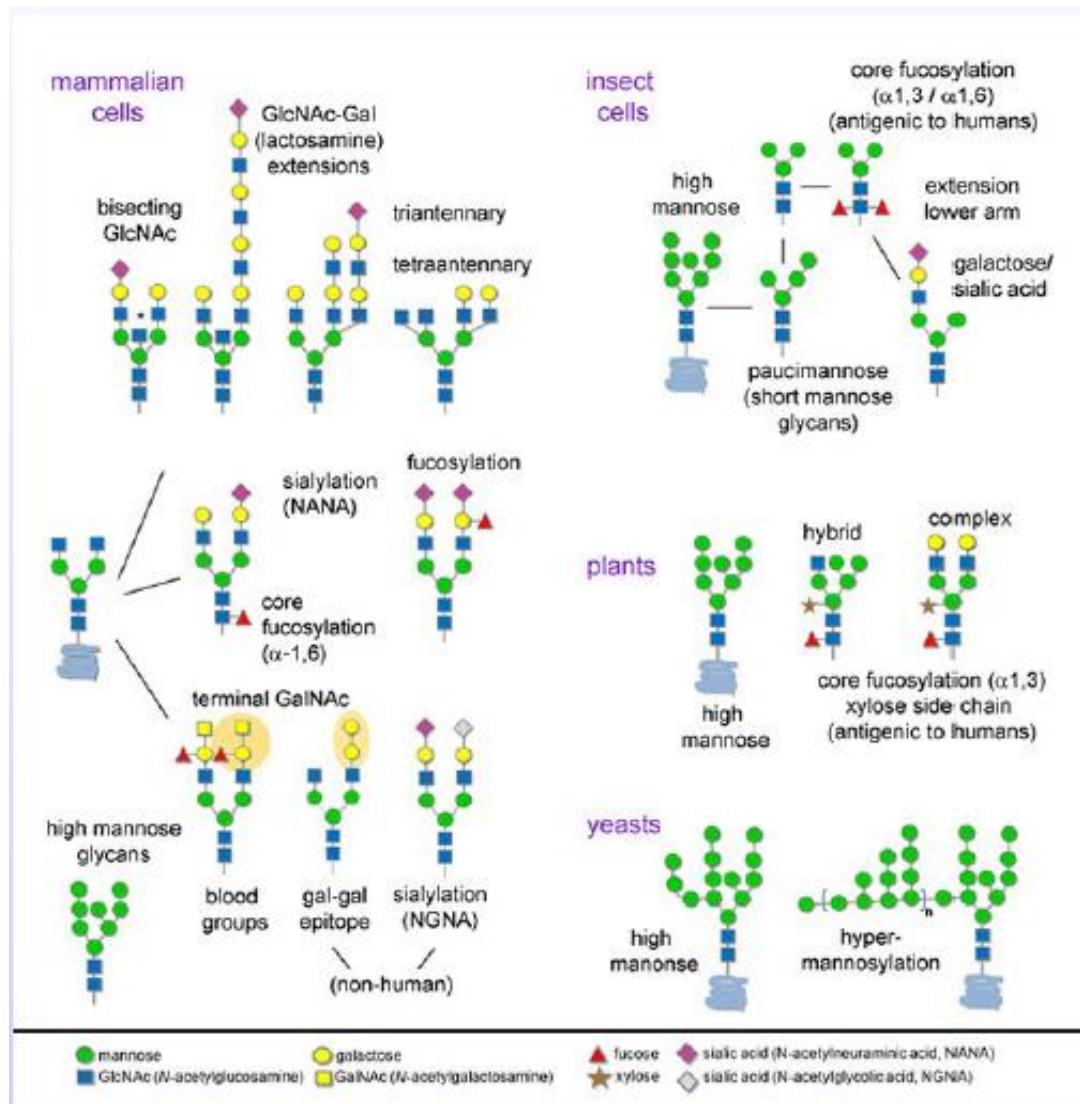
Richiede condizioni ossidanti

# Glicosilazione delle proteine.

## Ruolo della glicocomponente

- La glicosilazione può assistere il **folding** della catena polipeptidica attraverso effetti sulla struttura secondaria
- Le catene oligosaccaridiche possono contribuire alla **stabilità** di una proteina aumentandone la solubilità, schermando regioni idrofobiche, proteggendo dalle proteasi o favorendo interazioni intra-catena
- La glicosilazione può indirizzare una proteina alla sua destinazione finale (**trafficking/sorting**)
- 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

# La glicosilazione è una modifica post-sintetica organismo-specifica



# Organismo ospite

- **Procarioti**

- Batteri (*Escherichia coli*, *Bacillus sp.*)

- **Eucarioti**

- Lievito (*Saccharomyces cerevisiae*, *Pichia pastoris*, *Kluyveromyces lactis*)

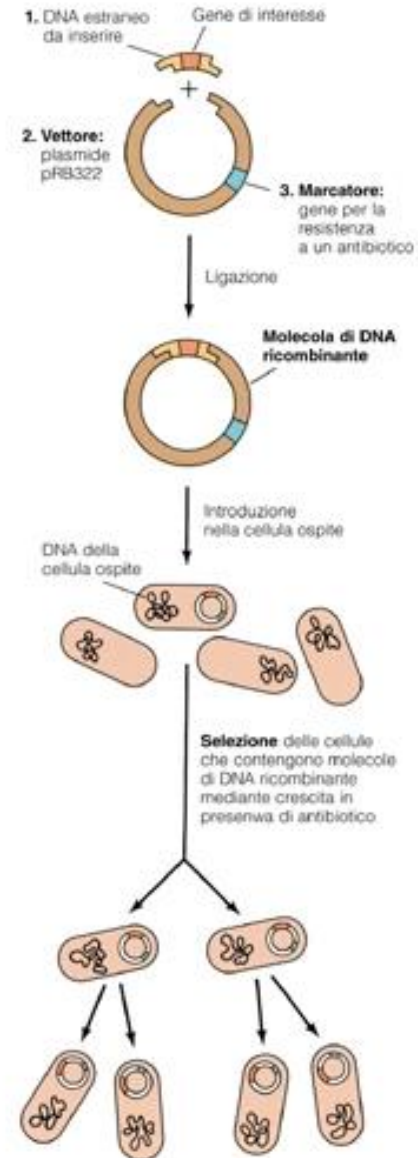
- Cellule di mammifero (CHO, BHK, HEK-293)

- Cellule di insetto (Sf9)

- Piante e animali transgenici

# Escherichia coli

- Crescita rapida (ca 30 min/raddoppio)
- Sistema economico
- Caratterizzazione dettagliata dei meccanismi molecolari dei processi di trascrizione e traduzione
- Disponibilità di ceppi mutanti
- Trasformazione semplice ( $\text{CaCl}_2$ , elettroporazione)
- Elevati livelli di espressione da plasmidi multicopia con promotori forti inducibili (lac, tac) e costitutivi (T7)
- Selezione mediante resistenza ad antibiotici

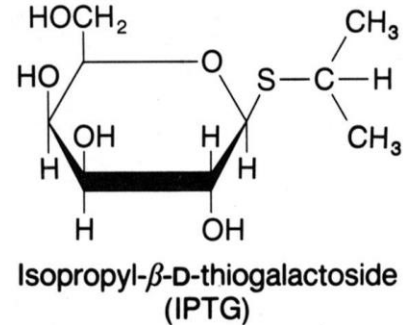




<b>Ceppo <i>E. coli</i></b>	<b>Caratteristiche</b>
BL21(DE3)	Esprime la RNA polimerasi del fago T7 a seguito di induzione con IPTG. È difettivo delle proteasi <i>lon</i> e <i>ompT</i>
BL21(DE3)-pLysS	Come le BL21(DE3) esprime la T7 RNA polimerasi. Esprime dal plasmide pLysS il lisozima T7. Adatto per l'espressione di proteine tossiche
Origami	Ha mutazioni sui geni <i>trxB</i> e <i>gor</i> . Facilita la formazione dei ponti disolfuro citoplasmatici
Origami-pLysS	Come il precedente con l'aggiunta del sistema pLysS per il controllo trascrizionale
Rosetta	Sovraesprime i tRNA per i codoni rari. Facilita l'espressione di proteine eucariotiche
Rosetta-gami-pLysS	Come il precedente con l'aggiunta del sistema pLysS per il controllo trascrizionale
BL21-codon plus	Sovraesprime i geni dei tRNA più rari per arginina (AGA/AGG), isoleucina (AUA) e leucina (CUA)
C41 e C43 (DE3)	Adatti per l'espressione di proteine tossiche di ogni origine

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

- Promotore lac (induzione con IPTG, un analogo del galattosio)



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

## *E. coli*: limitazioni?

- Espressione preferenzialmente intracellulare
- Aggregazione e corpi di inclusione
- Modifiche post-sintetiche?

# Lievito: *Saccharomyces cerevisiae*

- Crescita rapida (ca 90 min/raddoppio)
- Sistema economico
- Caratterizzazione dettagliata dei meccanismi molecolari dei processi di trascrizione e traduzione
- Disponibilità di ceppi mutanti (auxotrofie, proteasi)
- Trasformazione semplice (LiCl, elettroporazione)
- Plasmidi episomali o integrativi con promotori forti inducibili (GAL1, PHO5) o costitutivi (GAP, TPI, PGK)
- Selezione mediante complementazione di una auxotrofia
- Espressione intracellulare o secreta (sequenza segnale  $\alpha$ -factor)
- Modifiche post-sintetiche

*S. cerevisiae* si trasforma con un frequenza di almeno di 3 ordini di grandezza inferiore a quella di *E. coli*

può mantenere plasmidi che si replicano ma il numero di copie è molto inferiore a quello che si può ottenere in *E. coli* (max. 50)

Le procedure di clonaggio possono essere difficoltose, quindi *E. coli* viene utilizzato come intermedio nella produzione dei plasmidi ricombinanti, attraverso l'uso di vettori shuttle

- I plasmidi vengono assemblati *in vitro*
- trasformati, manipolati, modificati e propagati in *E. coli*
- e poi trasformati in lievito

## MARCATORI DI SELEZIONE

I lieviti sono naturalmente resistenti agli antibiotici

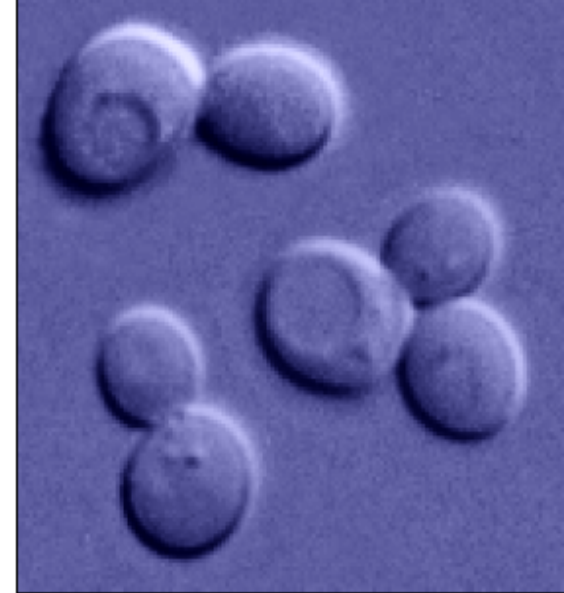


i geni di resistenza non sono marcatori idonei  
per la selezione dei trasformanti

In lievito si usano marcatori genici  
basati su auxotrofie

*HIS3, URA3, TRP1, LEU2, LYS2, ADE2*

Alcuni di questi marcatori (*URA3, LYS2*)  
hanno il vantaggio di poter essere usati per



**SELEZIONI  
POSITIVE**

basate sulla complementazione auxotrofica  
delle corrispondenti mutazioni

**SELEZIONI  
NEGATIVE**

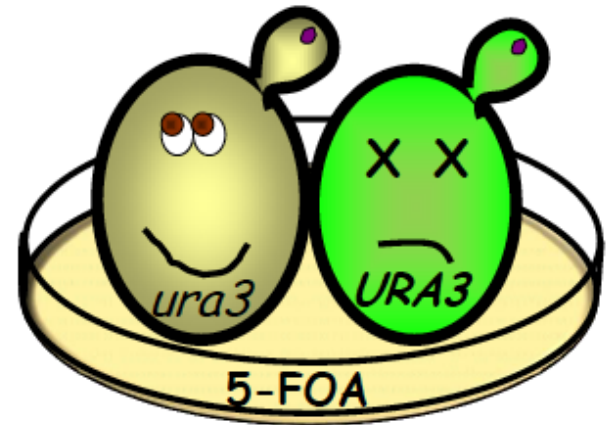
Basate su inibitori specifici che prevengono la crescita dei ceppi  
prototrofici ma non quella dei mutanti auxotrofi

## SELEZIONE NEGATIVA URA3

*URA3* codifica la orotidina-5'fosfato decarbossilasi, un enzima richiesto per la biosintesi dell'uracile

I mutanti *ura3-* (o *ura5-*) possono essere selezionati su terreni contenenti 5-FOA (acido 5-fluoroorotico)

Il prodotto di *URA3* converte il 5-FOA in un prodotto tossico che uccide le cellule wild type che lo esprimono



5-FOA è molto discriminante  
e può essere usato

per produrre marcatori *ura3*  
per mutazione in ceppi aploidi  
Per eliminare plasmidi con  
URA-3 (YCp, YEp e YIp)

Le piccole dimensioni del gene e la facilità della selezione negativa hanno fatto di URA-3 uno dei marcatori in assoluto più popolari per i lieviti

Nell'allele impiegato *ura3-52* è inserito un trasposone Ty1, questo evita l'integrazione di plasmidi Yip-URA3 nel locus *URA3* nella maggior parte dei ceppi di lievito

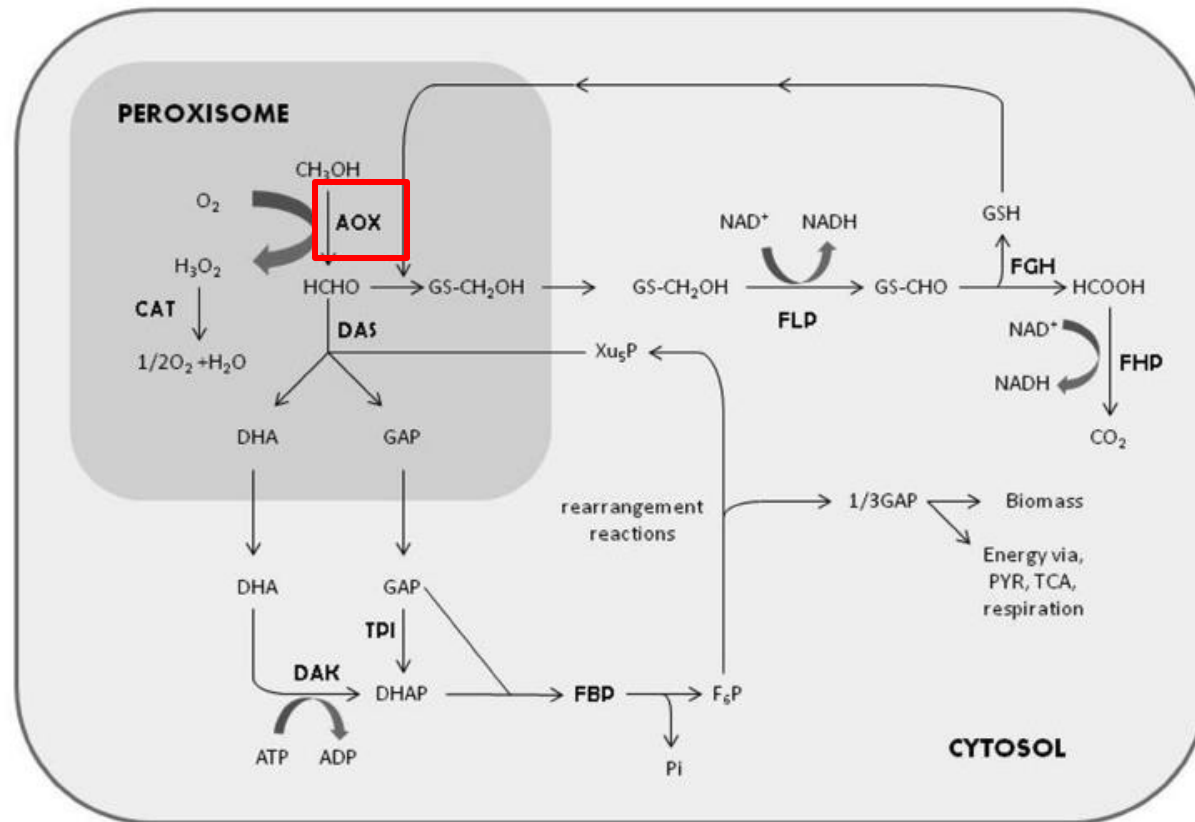
# Lievito: *Pichia pastoris* e *Kluyveromyces lactis*

- Crescite ad alta densità
- *P. pastoris* e *K. lactis* sono considerati GRAS (Generally Regarded As Safe)
- Sistemi di esporto delle proteine più efficienti
- Glicosilazione più simile a quella di mammifero
- Promotori inducibili
  - *P. pastoris* **AOX1** (metanolo), FLD (metanolo o metilamina), CUP1 (rame)
  - *K. lactis* LAC4 (galattosio, lattosio), PHO5 (fosfato)
- Promotori costitutivi
  - *P. pastoris* GAP, YPT1
  - *K. lactis* PGK
- Disponibilità di ceppi e vettori più limitata





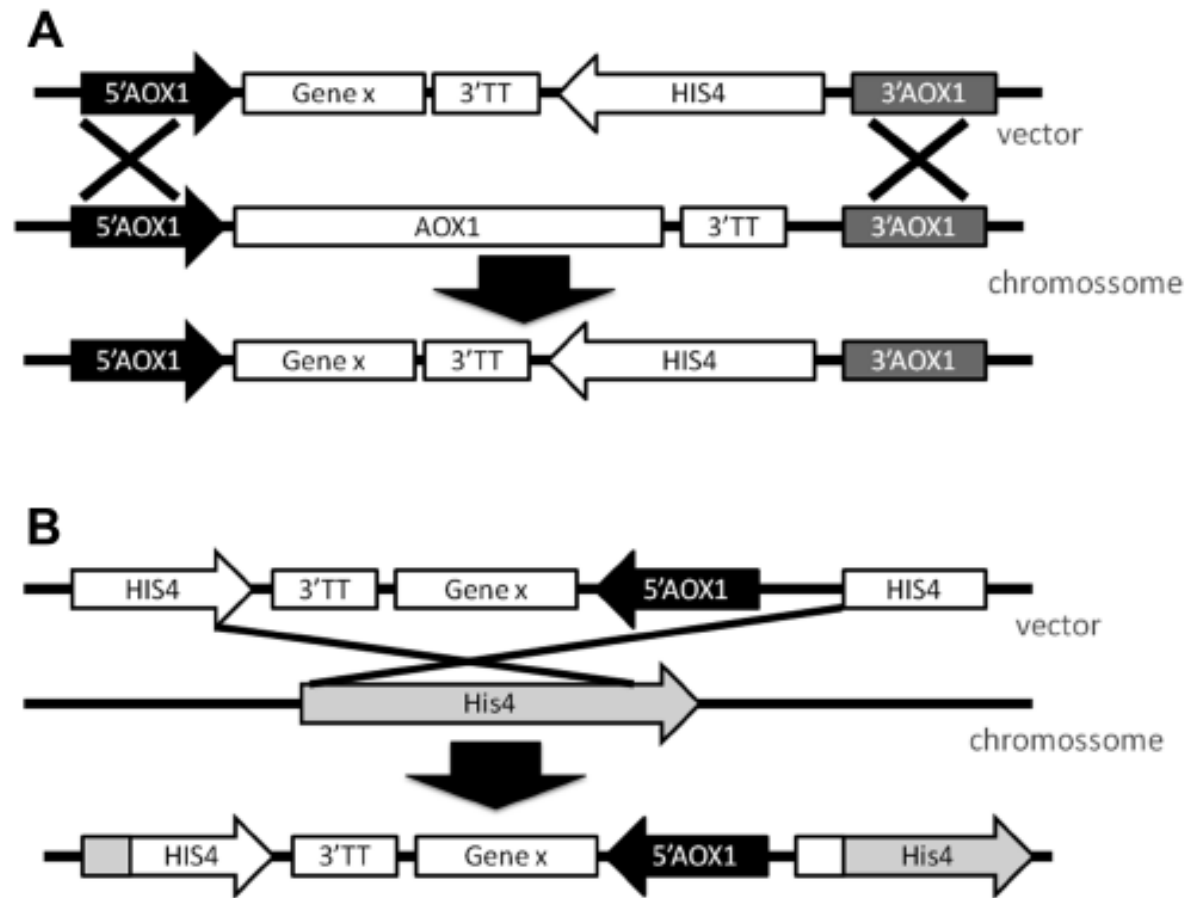
# Metabolismo del metanolo in *P. pastoris*



**Fig. 2.** Methanol metabolic pathway can be split into three steps.

1. Methanol is oxidized to formaldehyde by alcohol oxidase, alcohol dehydrogenase, or the catalase peroxidative pathway. 2. Formaldehyde is rapidly oxidized to formic acid mainly by formaldehyde dehydrogenase. Catalase and aldehyde dehydrogenase can also mediate this step. 3. Formic acid, which is the principal toxic agent in methanol poisoning, is metabolized by formate dehydrogenase. The catalase peroxidative pathway can also oxidize formic acid to  $\text{CO}_2$ .

# Integrazione del vettore nel genoma di *P. pastoris* per ricombinazione omologa

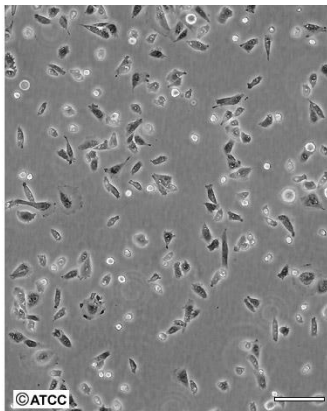


**Fig. 1.** Integration into *P. pastoris* genome by (A) gene replacement, and (B) gene insertion at his4.

# Cellule di mammifero

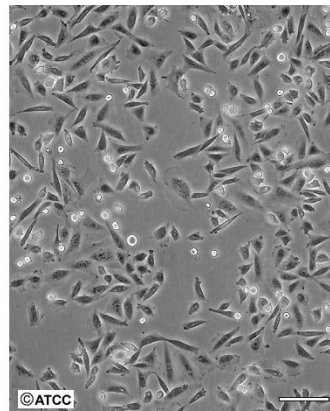
- Crescita lenta (ca 24 ore/raddoppio)
- Sistemi molto complessi e costosi
- Plasmidi episomali o integrativi con promotori forti inducibili (fitoestrogeni, TetOn) o costitutivi (CMV)
- Trasfezione transiente o stabile (lipocomplessi, elettroporazione)
- Trasduzione con vettori derivati da virus
- Selezione mediante resistenza ad antibiotici o vantaggio selettivo (diidrofolato reduttasi, glutamina sintasi)
- **Folding e modifiche post-sintetiche generalmente corrette**

ATCC Number: **CCL-61**  
Designation: **CHO-K1**



Low Density

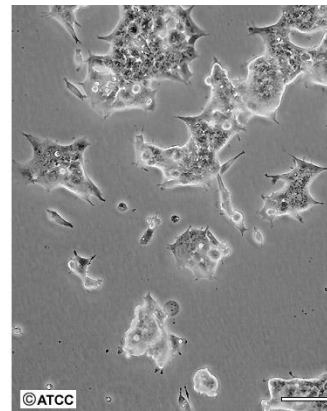
Scale Bar = 100µm



High Density

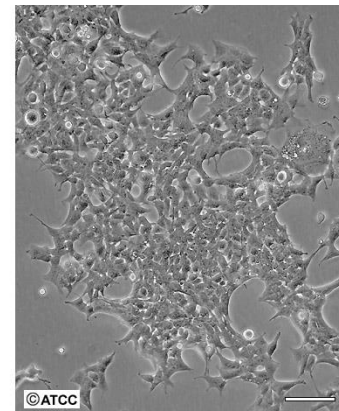
Scale Bar = 100µm

ATCC Number: **CRL-1573**  
Designation: **293**



Low Density

Scale Bar = 100µm

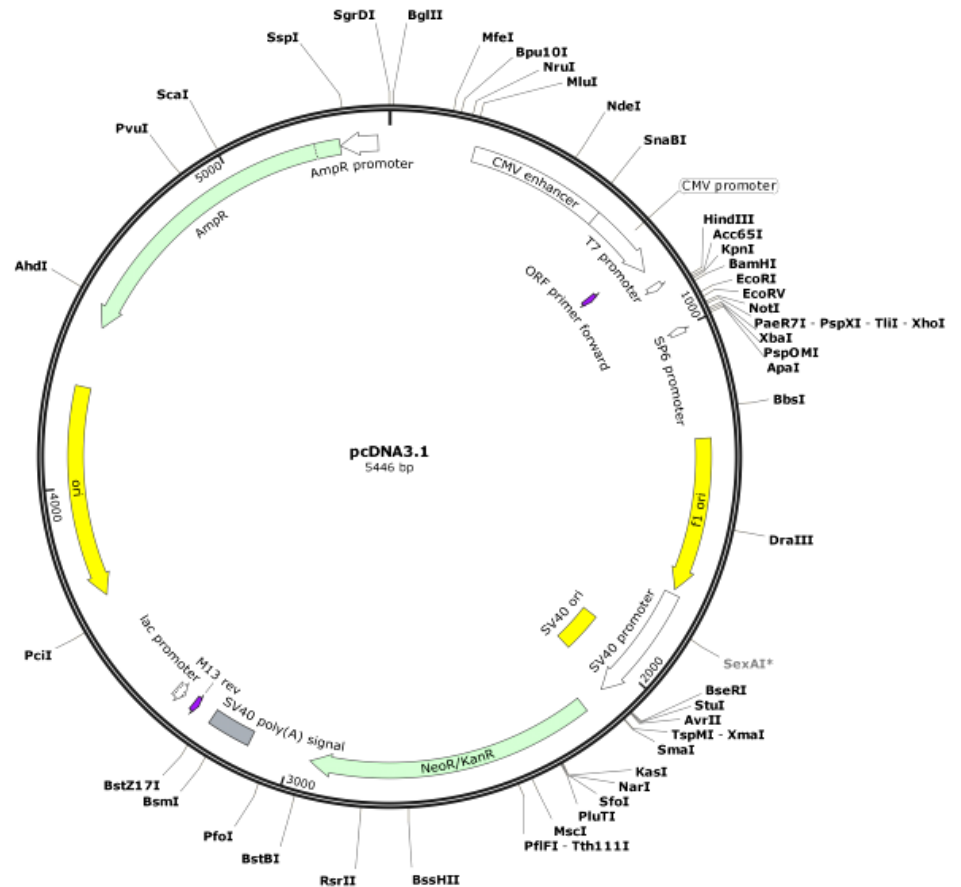


High Density

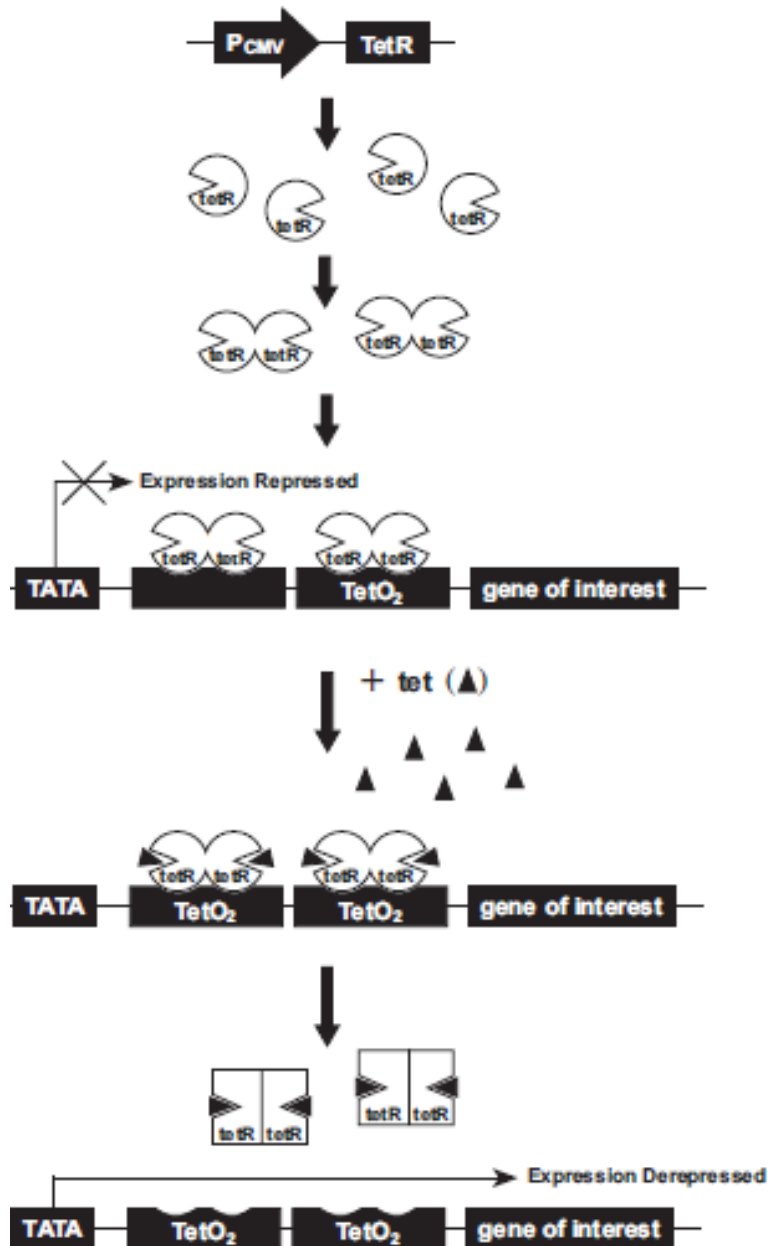
Scale Bar = 100µm

# Espressione in cellule di mammifero

- I vettori plasmidici sono più adatti per l'espressione di piccole quantità di proteina, in genere per studi che non richiedono l'isolamento della proteina

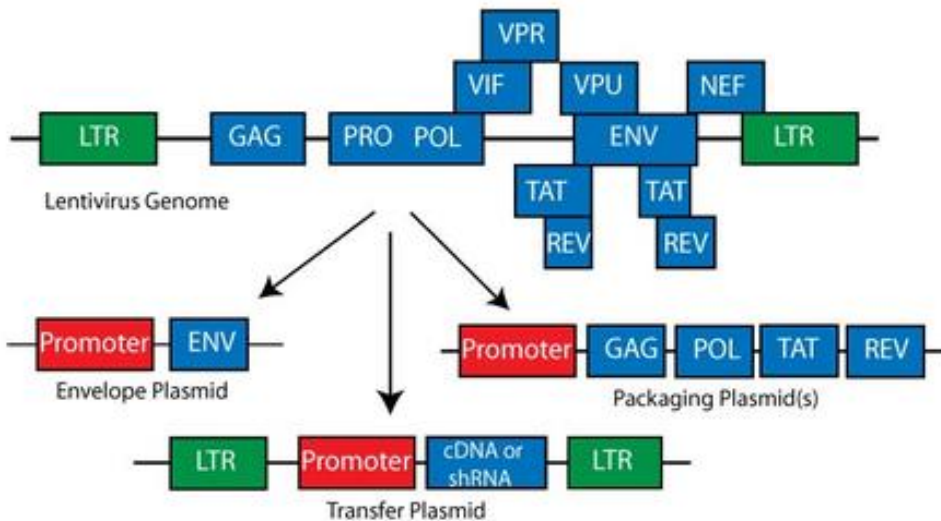
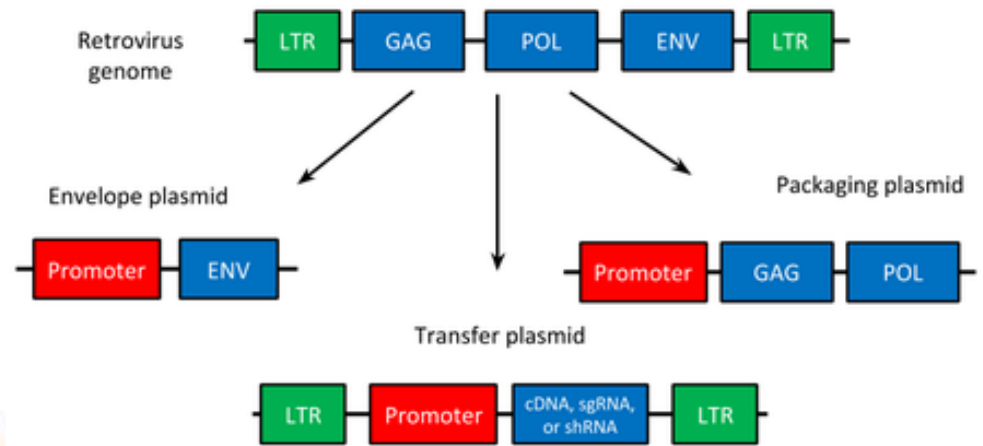


# Sistema inducibile TetOn



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, o dell'analogo doxyciclina, TetR non si lega al DNA e permette la trascrizione del gene di interesse.

# Sistemi basati su virus: lentivirus e retrovirus



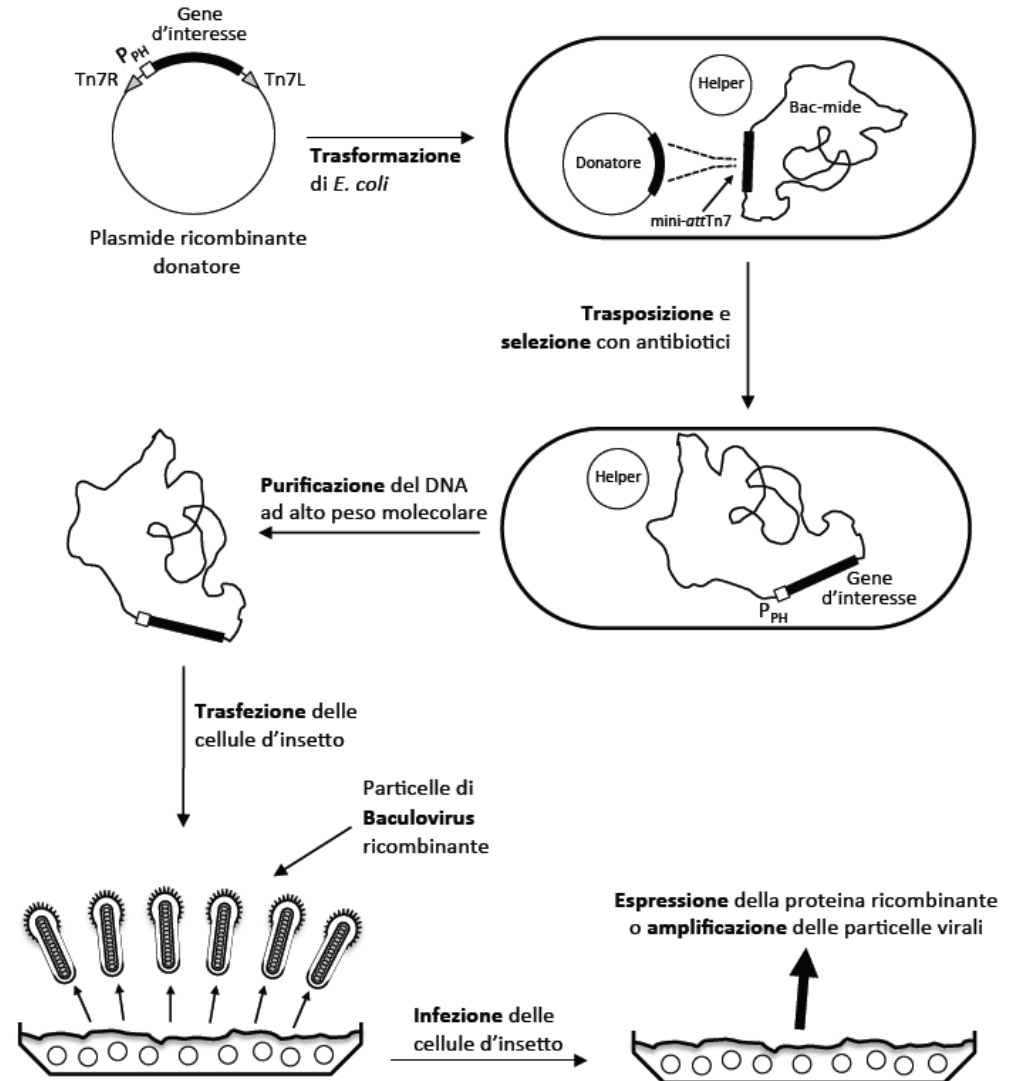
# Confronto di diversi sistemi basati su virus per l'espressione in cellule di mammifero



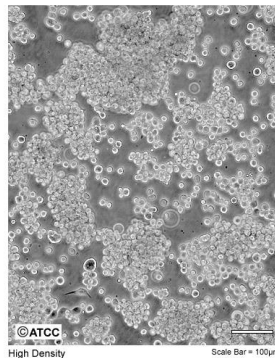
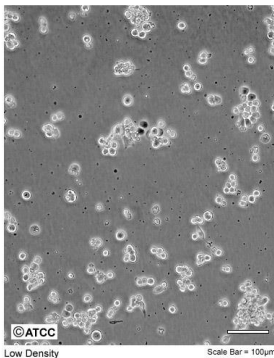
	ADENOVIRUS	AAV	$\gamma$ -RETROVIRUS	LENTIVIRUS
SIZE	~90-100 nm	~25 nm	~80-100 nm	~80-100 nm
GENOME	dsDNA	ssDNA	ssRNA	ssRNA
PACKAGING CAPACITY	~8 kb – 36 kb	~4.7 kb	10 kb	8 kb
TRANSDUCTION	Dividing and non-dividing cells	Dividing and non-dividing cells	Dividing cells	Dividing and non-dividing cells
TRANSDUCTION EFFICIENCY	High	Moderate	Moderate	Moderate
INTEGRATION	Non-integrating	Non-integrating	Integrating	Integrating
EXPRESSION	Transient	Transient or stable	Stable	Stable
BIOSAFETY LEVEL	BSL-2	BSL-1	BSL-2	BSL-2
IMMUNOGENICITY	High	Low	Moderate-High	Moderate-High
GENE THERAPY STRATEGY	<i>In vivo</i>	<i>In vivo</i>	<i>Ex vivo</i>	<i>Ex vivo</i>

# Cellule di insetto/baculovirus

- Cellule Sf9 infettate con baculovirus ricombinante
- Promotore forte della poliedrina
- Crescita lenta (ca 24 ore/raddoppio)
- Sistemi complessi e costosi
- **Folding e modifiche post-sintetiche generalmente corrette**



ATCC Number: CRL-1711  
Designation: Sf9





# 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 peptidici** o **proteici** fusi alla proteina (N- o C-terminali).  
Immobilizzazione di anticorpi contro i tag peptidici o di substrati dei tag proteici

# Tag peptidici e proteici

- Tag **peptidici** sono utili per purificazione e rivelazione
- Tag **proteici** aumentano i livelli di espressione e la **solubilità**, sono utili per purificazione e rivelazione
- Rimozione tag mediante inserimento di siti di taglio per proteasi nella sequenza linker tra la proteina e il tag

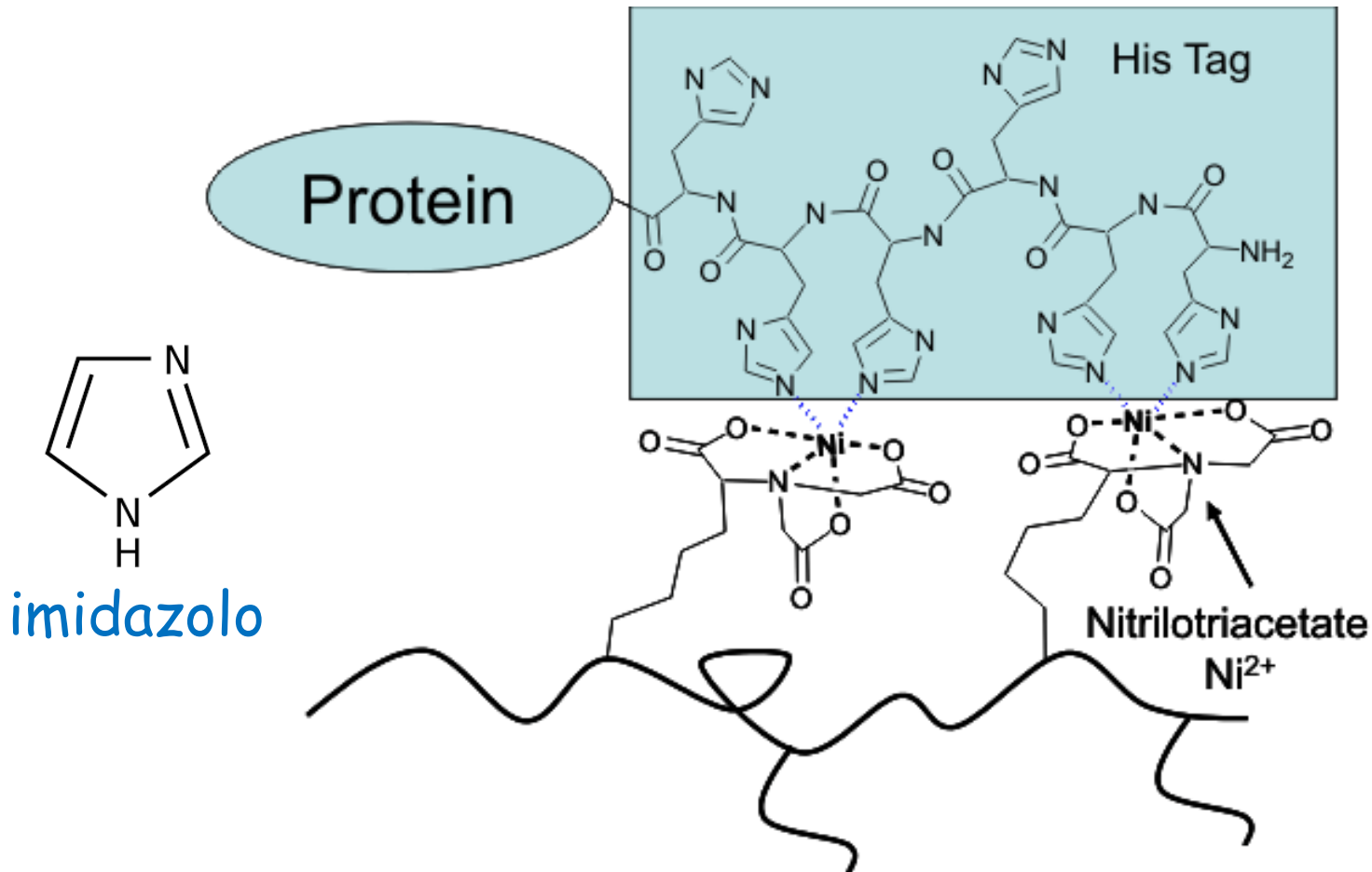
# Alcuni tag peptidici e proteici di uso comune

- **His tag** : HHHHHH 6-10 aa
- **FLAG** : DYKDDDDK 8 aa
- **Strep-tag II** : WSHPQFEK 8 aa
- **V5 tag**: GKIPNPLLGLDST 14 aa
- **HA tag**: YPYDVPDYA 9 aa
- **c-myc tag**: EQKLISEEDL 10 aa
  
- **Glutathione-S-transferase (GST) tag**: 211 aa
- **Maltose binding protein (MBP) tag**: 396 aa
- **Green fluorescent protein (GFP) tag**: 238 aa

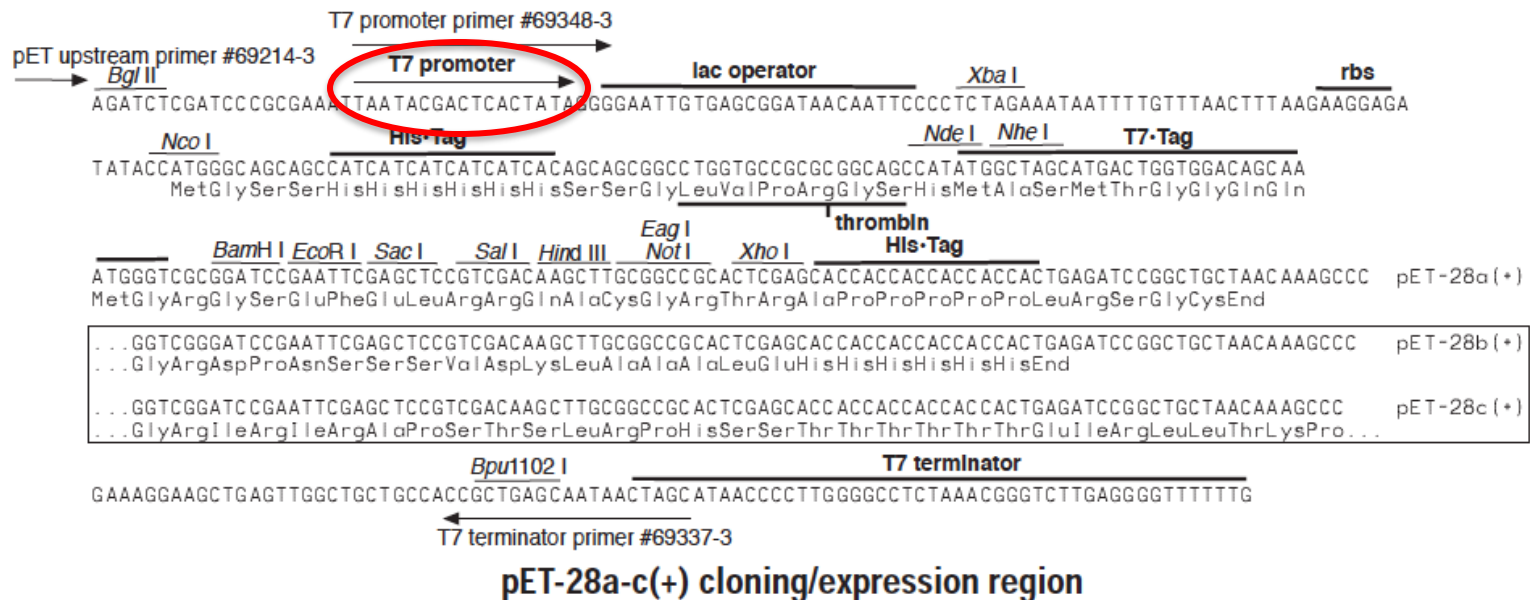
# Cromatografia di affinità per la purificazione delle proteine ricombinanti con tag

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, <del><math>\beta</math>-mercaptoethanol or cysteine</del>
Glutathione S-transferase	Glutathione	5–10 mM reduced glutathione
Maltose-binding protein	Cross-linked amylose	10 mM maltose

L'His-tag permette il legame a resine con metalli immobilizzati:  $\text{Ni}^{2+}$  o  $\text{Co}^{2+}$

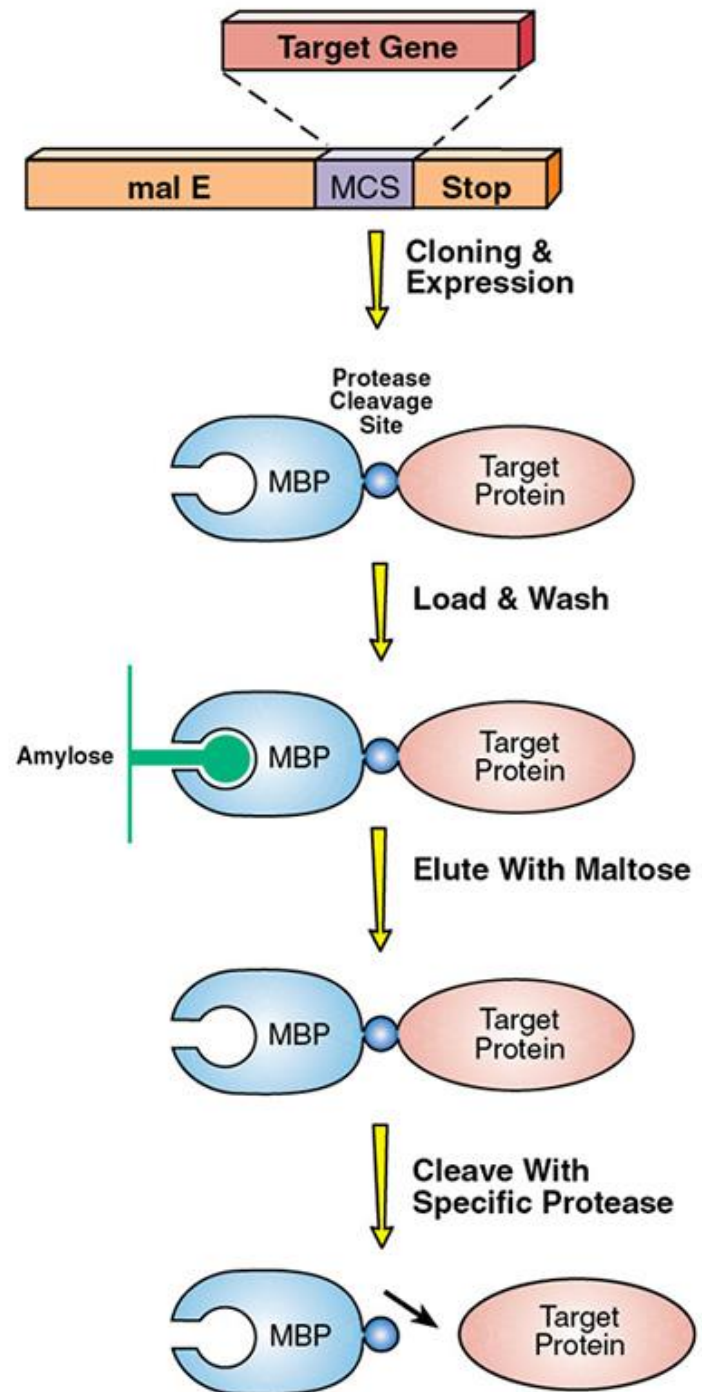
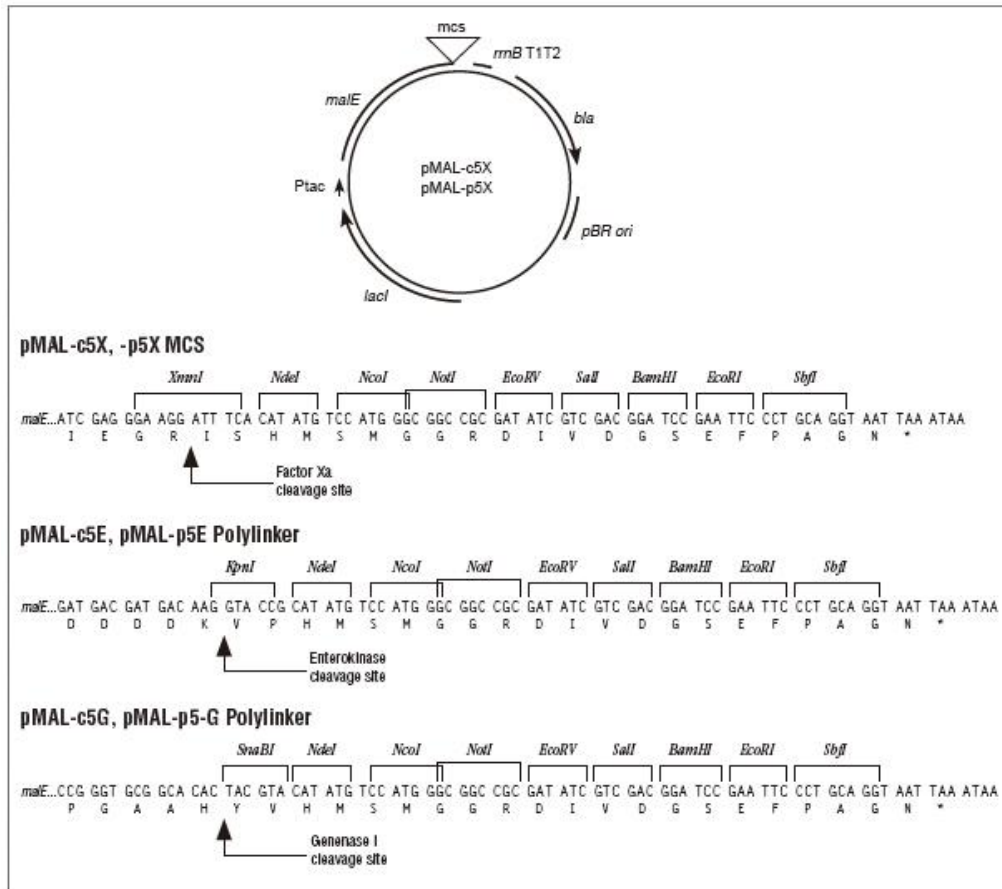


# Sistema pET per la produzione di proteine con His-tag in E. coli



- Promotore T7
- Sito di clonaggio che permette l'aggiunta di His-Tag N- o C-terminale
- Sito di taglio per la proteasi trombina per rimuovere l'His-Tag N-terminale
- Il ceppo di E. coli usato per l'espressione deve contenere la RNA polimerasi di T7

# Sistema per la produzione di proteine di fusione con MBP in E. coli

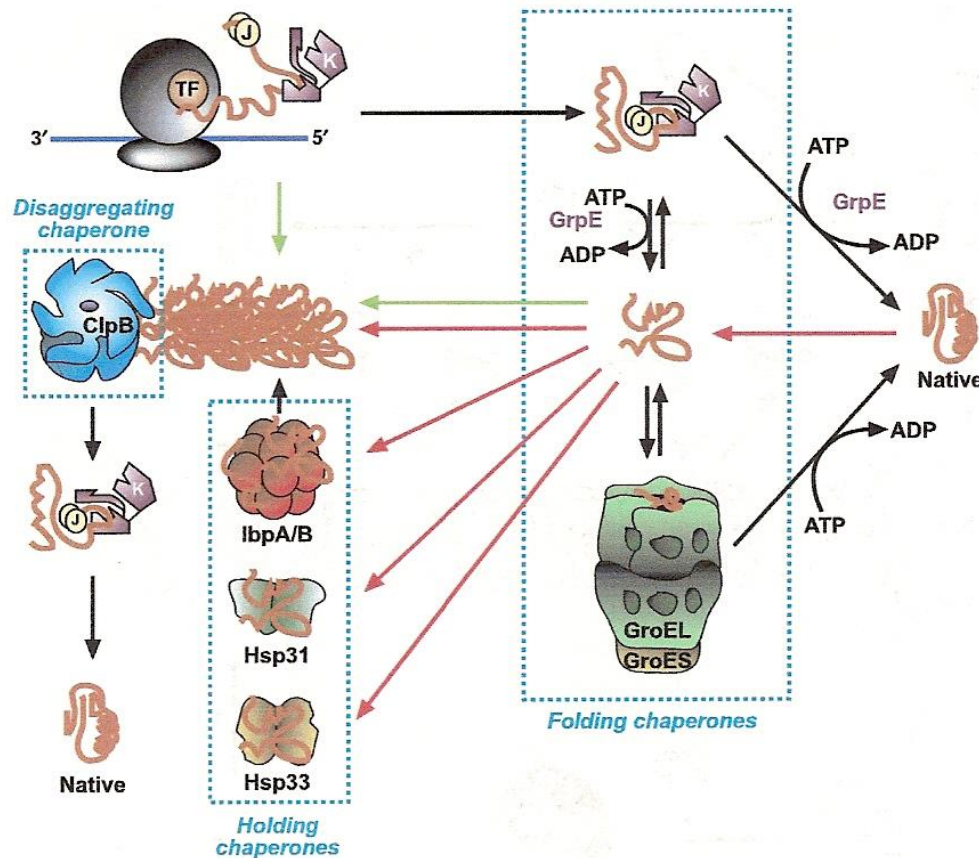


# Miglioramento dei livelli di espressione di una proteina ricombinante

- 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

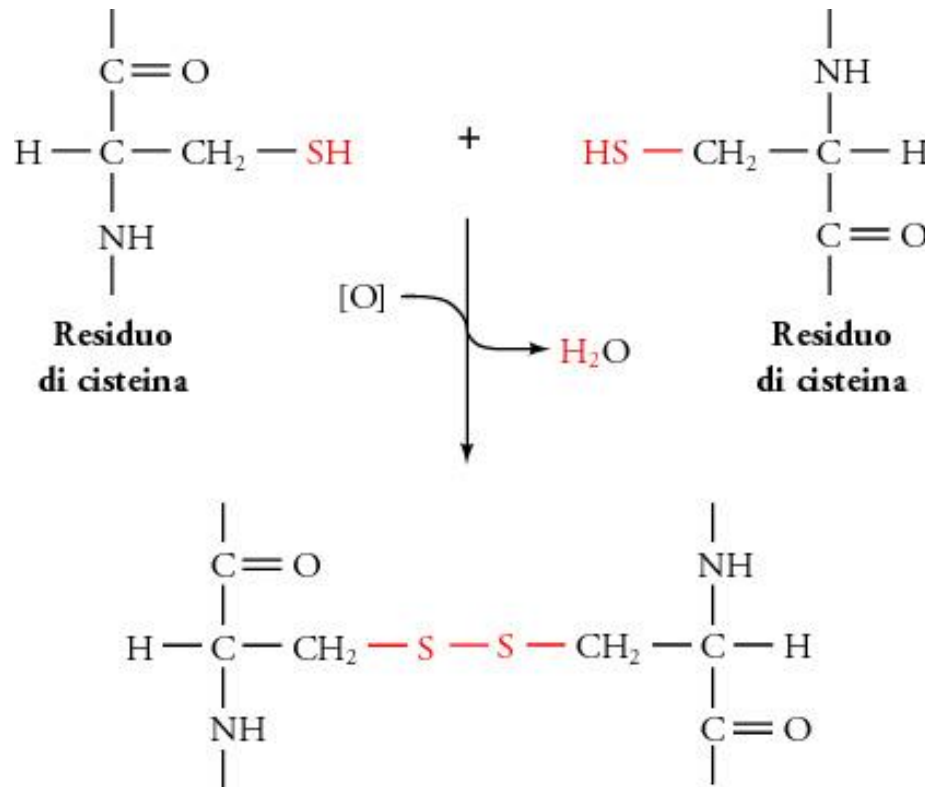


# Chaperon molecolari che assistono il folding delle proteine in *E. coli*



**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).

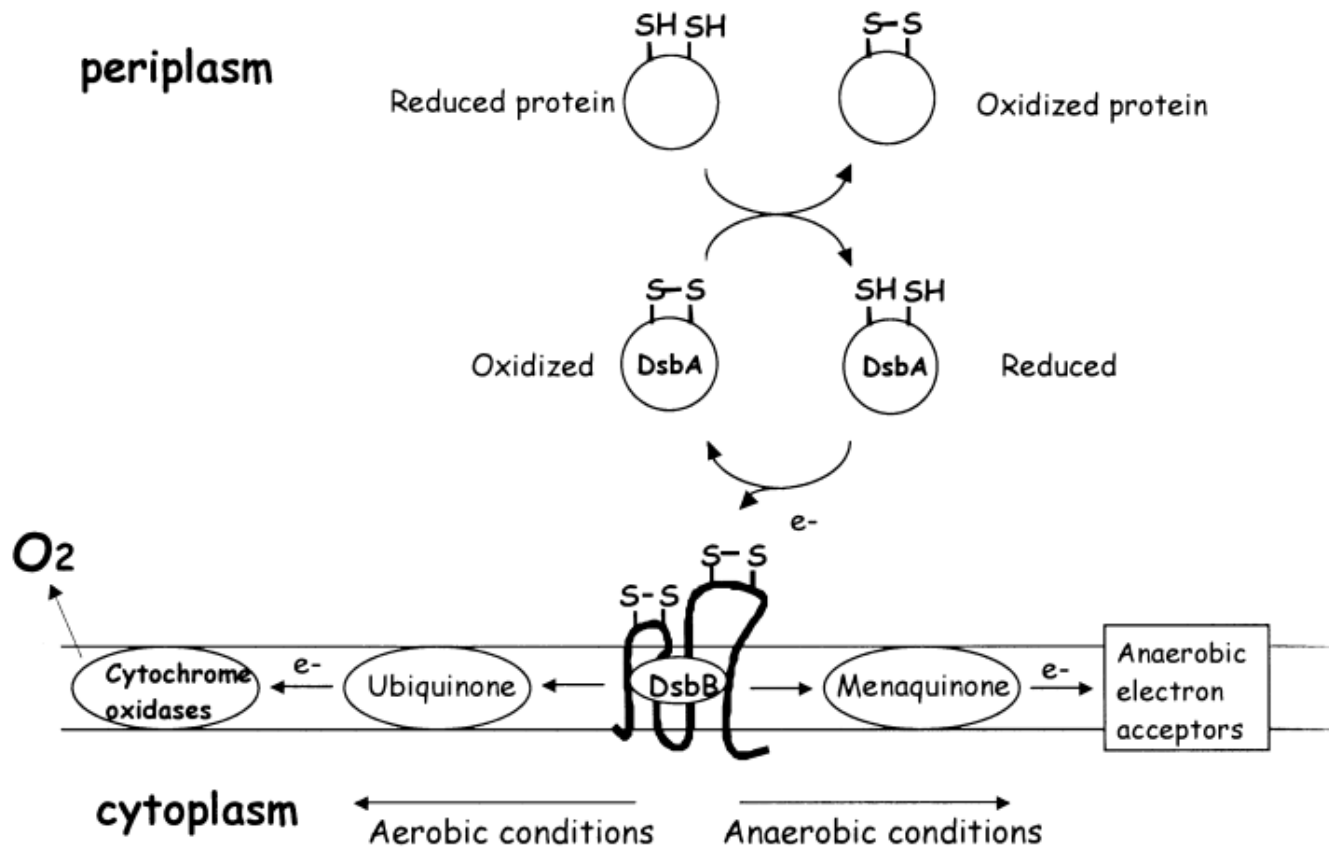
# Cisteina e ponti disolfuro



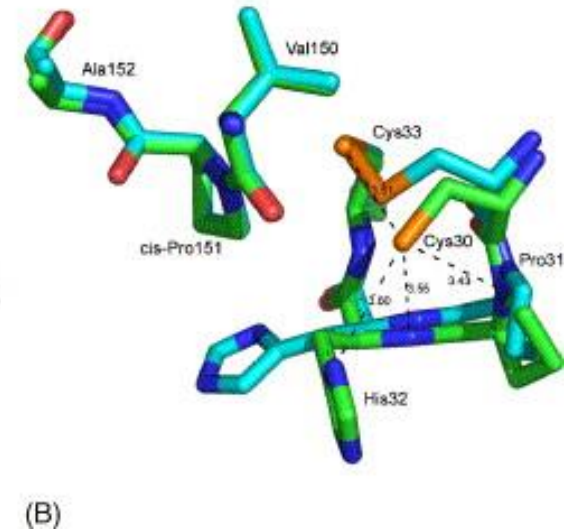
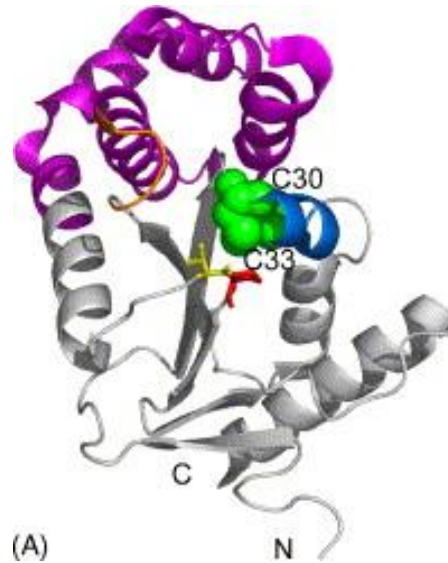
La formazione di ponti disolfuro avviene nel **periplasma** nei batteri e nel **reticolo endoplasmatico** negli eucarioti

Richiede condizioni ossidanti

# Sistema di formazione dei ponti disolfuro nel periplasma di *E. coli*



# Meccanismo d'azione di DsbA



Sito attivo **Cys**<sub>30</sub>Pro His **Cys**<sub>33</sub>

Maggiore stabilità nello stato ridotto

Alta reattività nello stato ossidato

Stabilizzazione stato ridotto che determina Cys30 con basso pKa, nello stato ridotto formazione:

- legame elettrostatico estremità positiva di una  $\alpha$ -elica
- legame idrogeno con His32.

DsbA ossidata ha una struttura flessibile che permette il legame di un gran numero di substrati

DsbA ridotta ha una struttura rigida che permette il legame solo di DsbB

# Isomerasi: DsbC / DsbG

NADPH

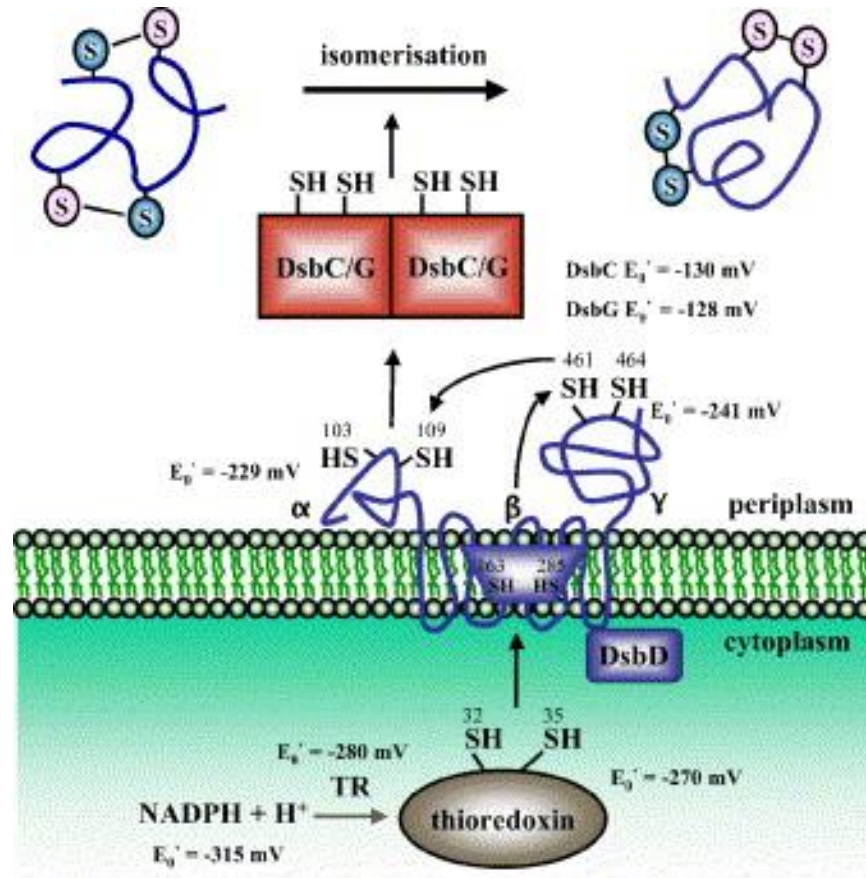
Tioredoxina reduttasi

Tioredoxina

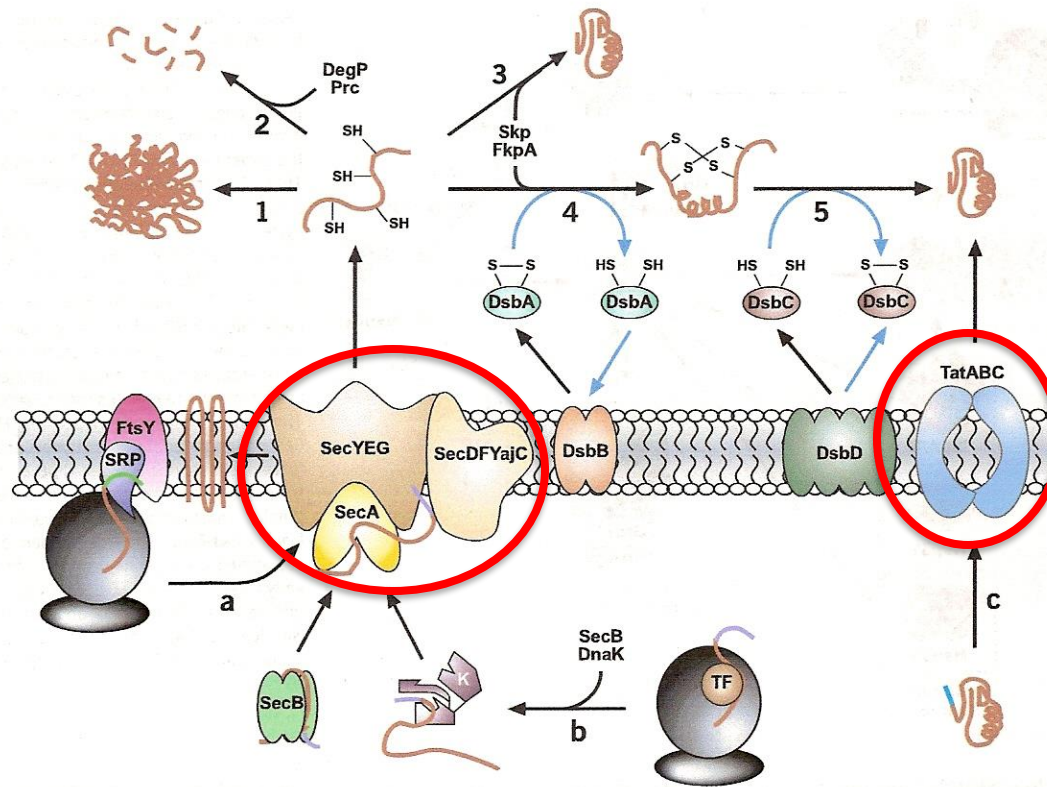
DsbD $\alpha$ -DsbD $\beta$ - DsbD $\gamma$

DsbC/DsbG

Rottura ponti disolfuro non corretti



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

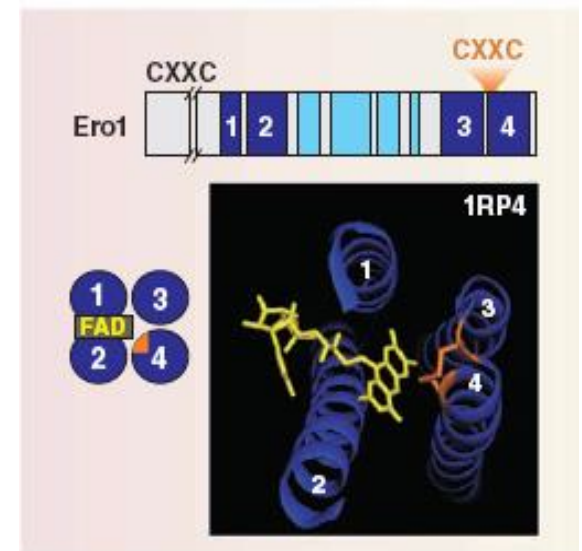
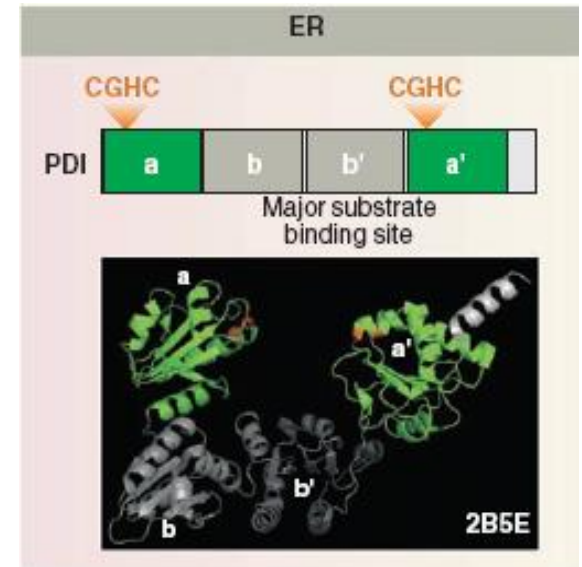
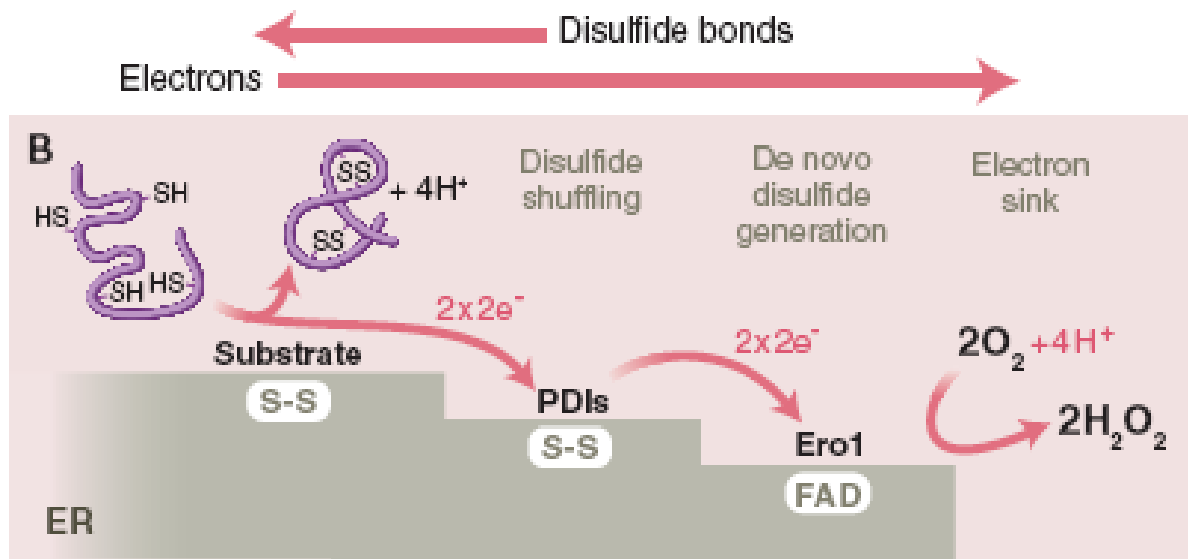


**Figure 2** Export and periplasmic folding pathways. Proteins destined for export can be translocated across the inner membrane in three different fashions. (a) Preproteins with highly nonpolar signal sequences (green) or transmembrane segments of inner membrane proteins are recognized by SRP which, 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
- Le catene oligosaccaridiche possono contribuire alla **stabilità** di una proteina aumentandone la solubilità, schermando regioni idrofobiche, proteggendo dalle proteasi o favorendo interazioni intra-catena
- La glicosilazione può indirizzare una proteina alla sua destinazione finale (**trafficking/sorting**)
- 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

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

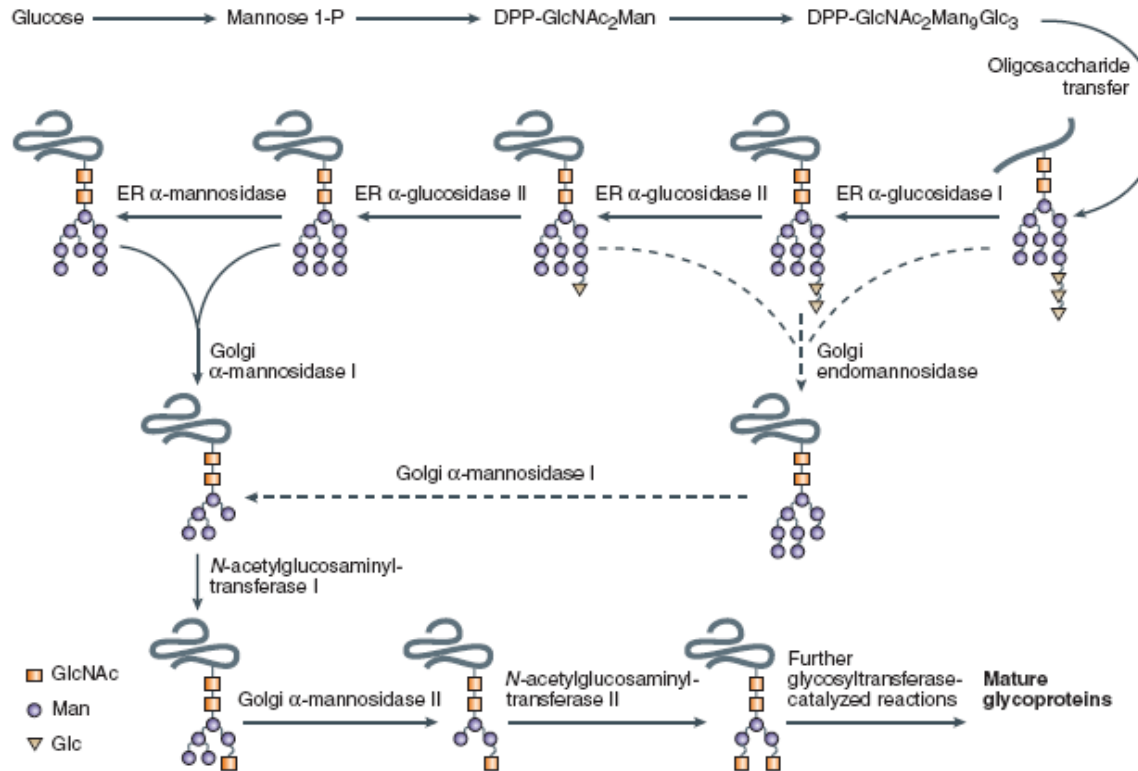
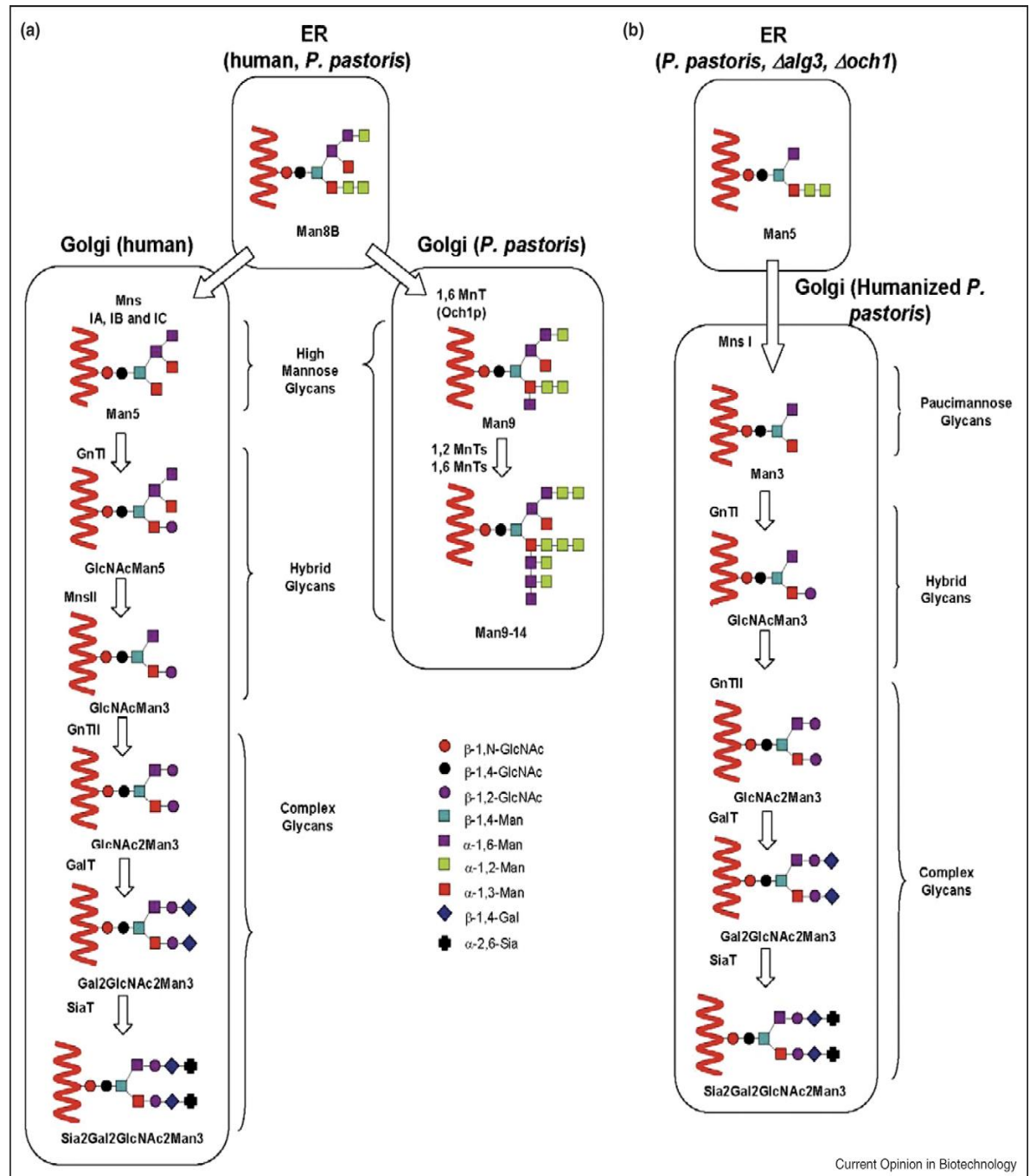


Figure 1 Overview of the biosynthesis of protein N-linked glycan side chains. As the nascent glycoprotein enters the ER, a preformed oligosaccharide, dolichol phosphate precursor (DPP) is attached cotranslationally to some asparagine residues that are part of the consensus sequence Asn-Xaa-Ser/Thr (where 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.)

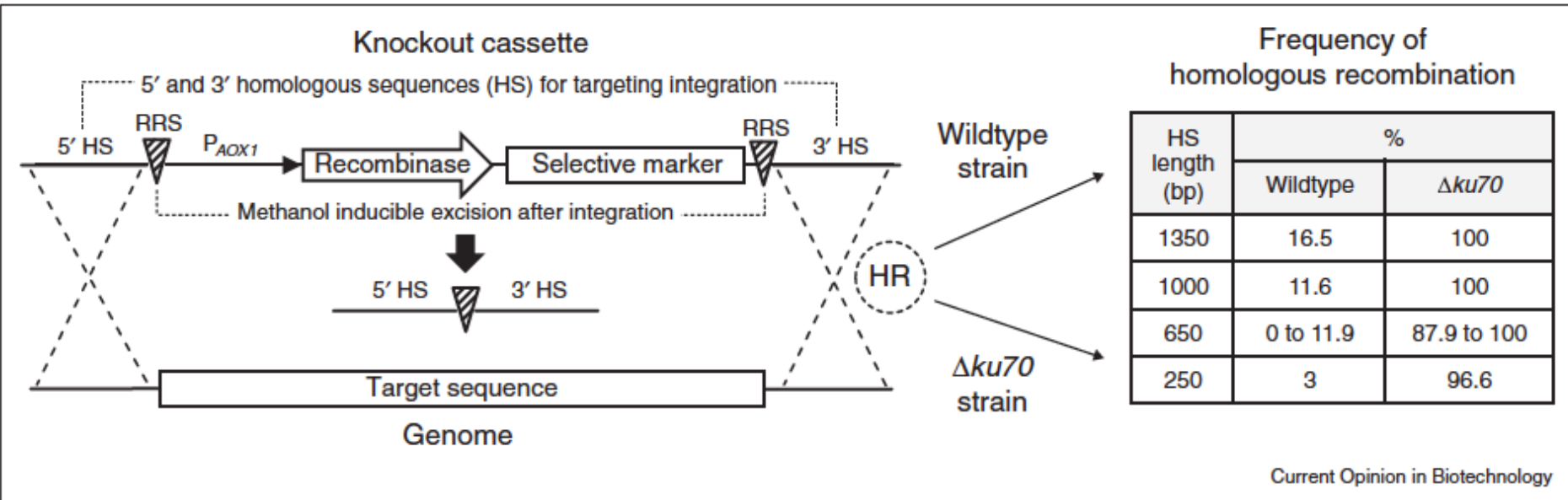
# Umanizzazione della glicosilazione del lievito *Pichia pastoris*

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.



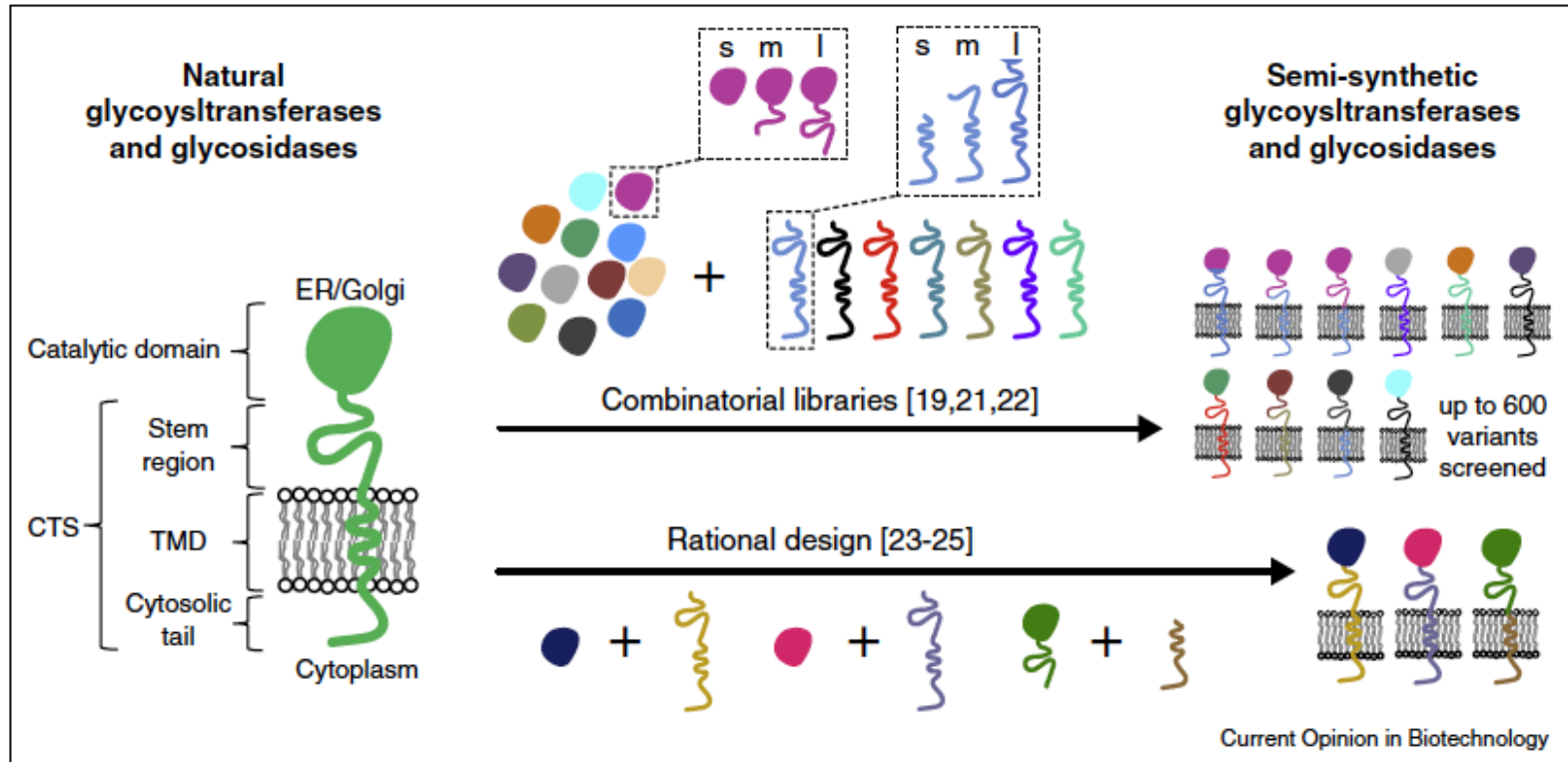
# Costruzione di ceppi di lievito mutanti

## Cassette per la rigenerazione di marker di selezione dopo inattivazione di un gene



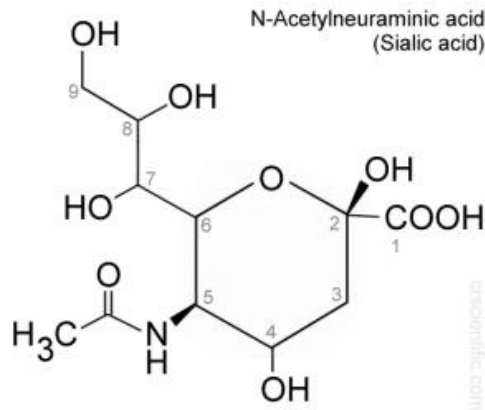
Recombinase based self-excisable knockout cassettes for marker regeneration (left side). Increased rates of homologous recombination in a *P. pastoris*  $\Delta ku70$  strain (right side). The knockout cassettes consist of a recombينase (Cre or FLP [48,49,50\*]) and a marker gene flanked by the respective recombينase recognition sites and are directed to the genome via the 5' and 3' homologous sequences to delete the desired target sequence. After integration via a double cross-over event, self-excision of the recombينase and the marker gene can be initiated by the expression of the recombينase from the methanol inducible  $AOX1$  promoter ( $P_{AOX1}$ ), leaving only the recombينase recognition site in the genome (notably Marx *et al.* [49] provided the recombينase transiently on a CEN/ARS plasmid). The initial integration in the genome is dependent on homologous recombination (HR). Exemplary frequencies of homologous recombination (in %) of the wildtype compared to the  $\Delta ku70$  strain are shown (right side). The length of the homologous sequence indicates the number of base pairs (bp) added on both sides of the cassette [50\*]. For 650 bp two different integration loci were tested, therefore two % values are given.

# Strategie per costruire glicosiltrasferasi ibride



Design strategies to create semi-synthetic glycosyltransferases and glycosidases for glycoengineering. On the left side, the general domain structure of glycosyltransferases and glycosidases is shown. These type II membrane proteins consist of an N-terminal cytosolic tail, a transmembrane domain (TMD), a stem region (these elements are referred to as CTS), and a C-terminal catalytic domain. In the middle and on the right side, design strategies for creating tailor-made enzymes with the **desired catalytic activity and the proper localization** in the sec pathway are shown. The combinatorial library approach involved the combination of large sets of catalytic domains with CTS fragments to fusion proteins, which were then screened for the desired activity [19,21,22]. Different lengths of the catalytic domains and the CTS fragments were tested (referred to as 's' for short, 'm' for medium, 'l' for long and shown exemplarily for one catalytic domain and one CTS). Rational approaches were also used to design these chimeric enzymes [23–25]. The schematic for the domain architecture and the combinatorial libraries is based on Czapinski *et al.* [20] and Nett *et al.* [26\*].

# Sintesi e incorporazione di acido sialico in glicoproteine in lievito



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-acetylneuraminate-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.

