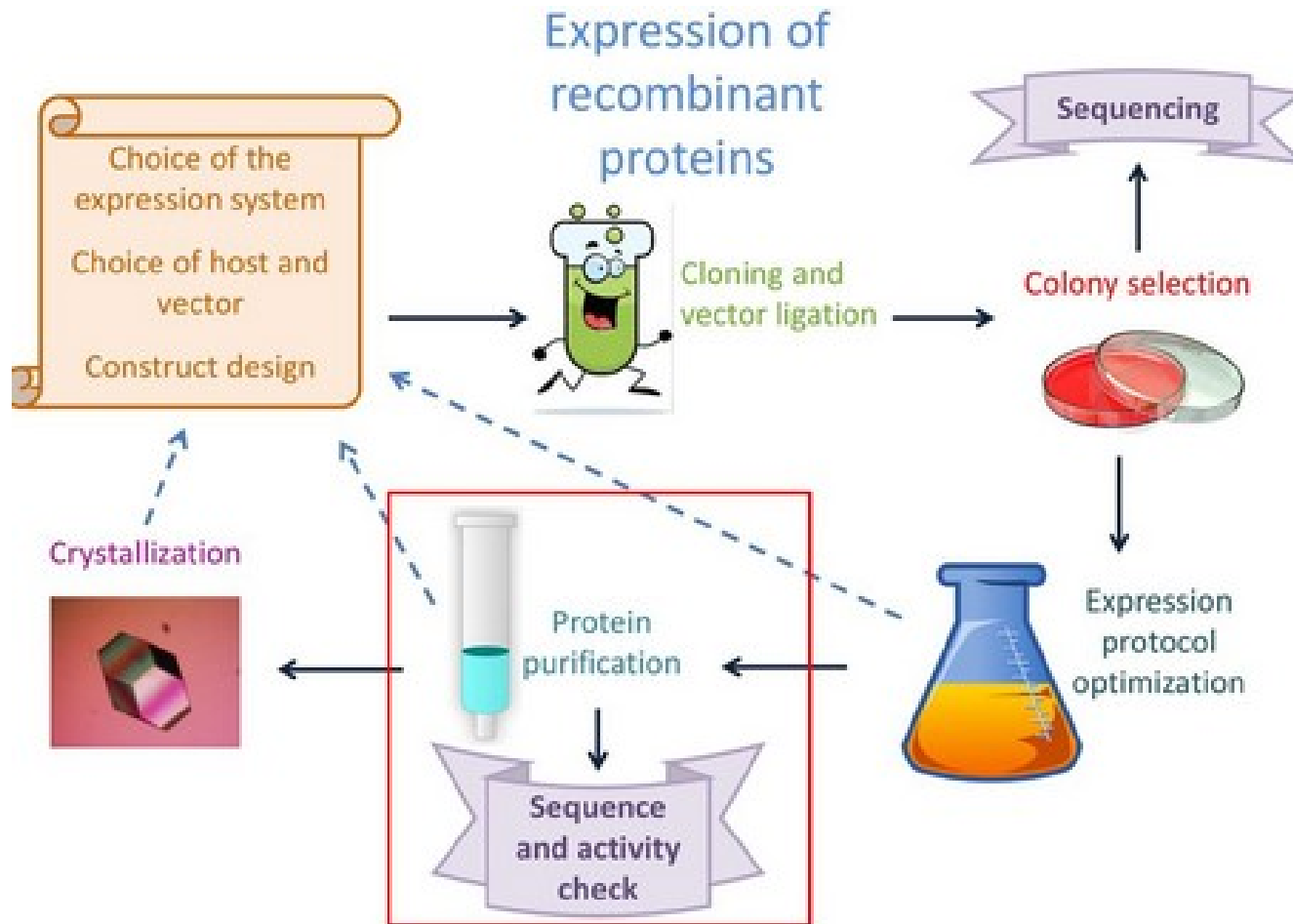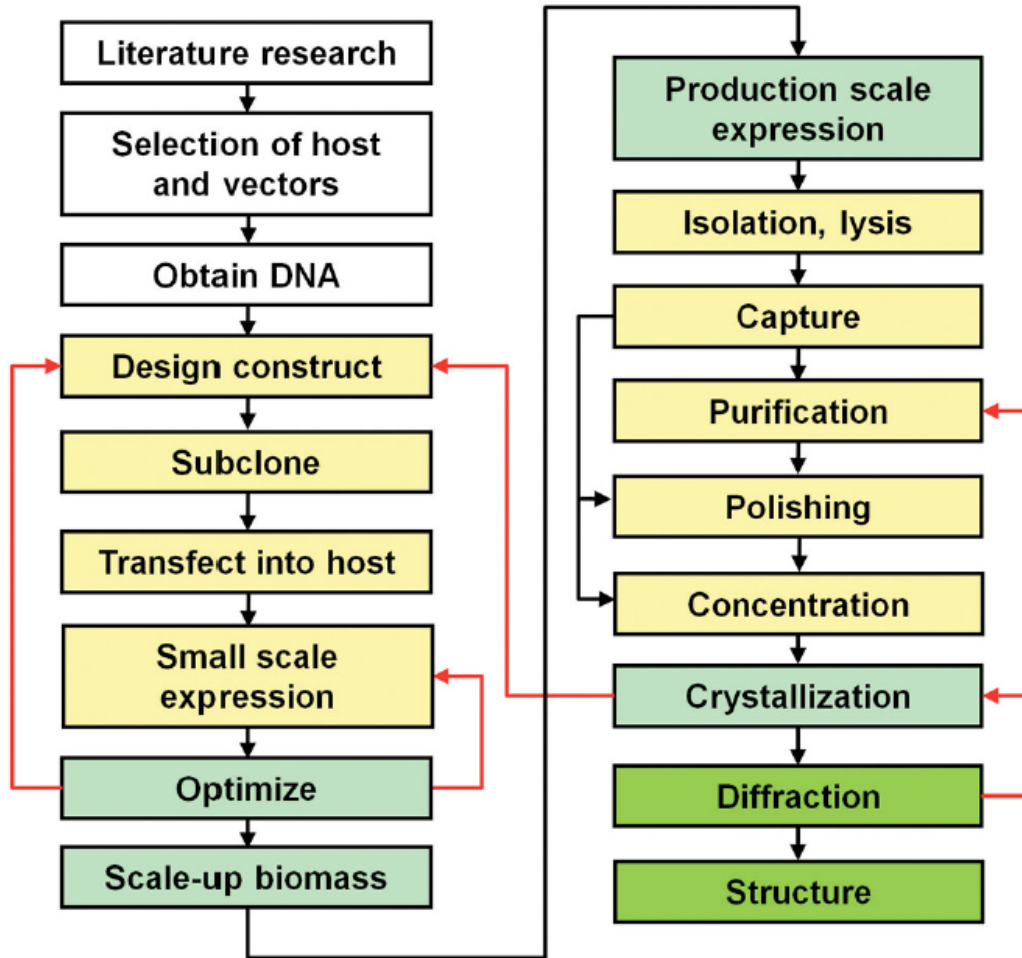# Proteine ricombinanti per studi strutturali:
espressione, purificazione e ottimizzazione

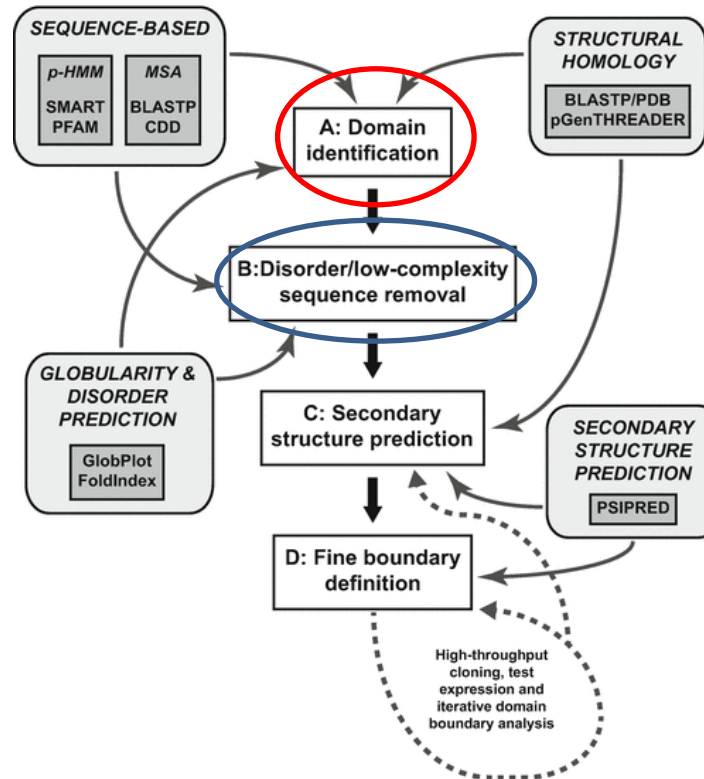# Espressione/purificazione di proteine ricombinanti per studi strutturali

Espressione/purificazione di proteine ricombinanti per studi strutturali flowchart

Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Analisi bioinformatica e Metodi sperimentali

1) Analisi della sequenza (Codon usage, Siti di proteolisi, sequenze di secrezione)
2) Definizione dei domain boundaries

Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Analisi bioinformatica: A)   disorder prediction:

1)   FOLDINDEX: https://fold.weizmann.ac.il/fldbin/findex

I =2.785<H>−|<R>|−1.151

con | <R> | carica netta media
               <H> idrofobicità media


```
Summary:
Number Disordered Regions:  4
Longest Disordered Region:  72
Number Disordered Residues: 147
Predicted disorder segment: [260]-[264] length:  5 score: -0.01 ± 0.01
Predicted disorder segment: [286]-[293] length:  8 score: -0.02 ± 0.02
Predicted disorder segment: [389]-[460] length: 72 score: -0.25 ± 0.10
Predicted disorder segment: [462]-[523] length: 62 score: -0.08 ± 0.05
```

```
  1 HWPSEPSEAV RDWENQLEAS MHSVLSDLHE AVPTVVGIPD GTAVVGRSFR
 51 VTIPTDLIAS SGDIIKVSAA GKEALPSWLH WDSQSHTLEG LPLDTDKGVH
101 YISVSATRLG ANGSHIPQTS SVFSIEVYPE DHSELQSVRT ASPDPGEVVS
151 SACAADEPVT VLTVILDADL TKMTPKQRID LLHRMRSFSE VELHNMKLVP
201 VVNNRLFDMS AFMAGPGNAK KVVENGALLS WKLGCSLNQN SVPDIHGVEA
251 PAREGAMSAQ LGYPVVGWHI ANKKPPLPKR VRRQIHATPT PVTAIGPPTT
301 AIQEPPSRIV PTPTSPAIAP PTETMAPPVR DPVPGKPTVT IRTRGAIIQT
351 PTLGPIQPTR VSEAGTTVPG QIRPTMTIPG YVEPTAVATP PTTTTKKPRV
401 STPKPATPST DSTTTTTRRP TKKPRTPRPV PRVTTKVSIT RLETASPPTR
451 IRTTTSGVPR GGEPNQRPEL KNHIDRVDAW VGTYFEVKIP SDTFYDHEDT
501 TTDKLKLTLK LREQQLVGEK SWVQFNSHSQ LMYGLPDSSH VGKHEYFMHA
551 TDKGGLSAVD AFEIHVHRRP QGDRAPARFK AKFVGDPALV LNDIHKKIAL
601 VKKLAFAFGD RNCSTITLQN ITRG
```

(Predicted disordered segment)

Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Analisi bioinformatica: A)   disorder prediction:

2)   GLOBPLOT: http://globplot.embl.de/
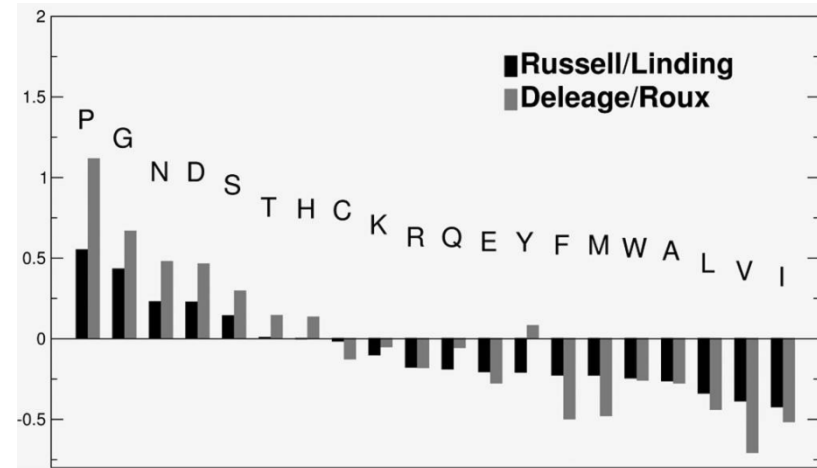
PROPENSIONE al disordine di ogni aa

$$P = RC - SS$$

RC random coil (loops, turns)
SS strutture secondarie regolari

Calcolata su un insieme non ridondante di proteine
da ciascun rappresentante da ciascuna superfamiglia
nel database **SCOP** (**Structural Classification of Proteins):**
2020-09-25 include **59,801** domini non redondanti in **682,669** strutture.

Disordered by Russell/Linding definition

```
>none_Disorder 2-6, 88-93, 111-117, 137-149, 237-244, 290-338, 352-377, 382-431, 449-469, 531-540, 616-624
hWPSEPseav rdwenqleas mhsvlsdlhe avptvvgipd gtavvgrsfr vtiptdlias sgdiikvsaa gkealpswlh wdsqshtLEG LPLdtdkgvh yisvsatrlg ANGSHIPqts svfsievype dhselqSVRT
ASPDPGEVVs sacaadepvt vltvildadl tkmtpkqrid llhrmrsfse velhnmklvp vvnnrlfdms afmagpgnak kvvengalls wklgcsLNQN SVPDihgvea paregamsaq lgypvvgwhi ankkpplpkr
vrrqihatpT PVTAIGPPTT AIQEPPSRIV PTPTSPAIAP PTETMAPPVR DPVPGKPTvt irtrgaiiqt pTLGPIQPTR VSEAGTTVPG QIRPTMTipg yVEPTAVATP PTTTTKKPRV STPKPATPST DSTTTTTRRP
TKKPRTPRPV Prvttkvsit rletasppTR IRTTTSGVPR GGEPNQRPEl knhidrvdaw vgtyfevkip sdtfydhedt ttdklkltlk lreqqlvgek swvqfnsnsq LMYGLPDSSH vgkheyfmha tdkgglsavd
afeihvhrrp qgdraparfk akfvgdpalv lndihkkial vkklafafgd rncstITLQN ITRG
```

Potential globular domains (GlobDoms) by Russell/Linding definition

```
>none_GlobDoms 1-289, 470-622
HWPSEPSEAV RDWENQLEAS MHSVLSDLHE AVPTVVGIPD GTAVVGRSFR VTIPTDLIAS SGDIIKVSAA GKEALPSWLH WDSQSHTLEG LPLDTDKGVH YISVSATRLG ANGSHIPQTS SVFSIEVYPE DHSELQSVRT
ASPDPGEVVS SACAADEPVT VLTVILDADL TKMTPKQRID LLHRMRSFSE VELHNMKLVP VVNNRLFDMS AFMAGPGNAK KVVENGALLS WKLGCSLNQN SVPDIHGVEA PAREGAMSAQ LGYPVVGWHI ANKKPPLPKR
VRRQIHATPt pvtaigpptt aiqeppsriv ptptspaiap ptetmappvr dpvpgkptvt irtrgaiiqt ptlgpiqptr vseagttvpg qirptmtipg yveptavatp pttttkkprv stpkpatpst dsttttttrrp
tkkprtprpv prvttkvsit rletaspptr irtttsgvpr ggepnqrpeL KNHIDRVDAW VGTYFEVKIP SDTFYDHEDT TTDKLKLTLK LREQQLVGEK SWVQFNSNSQ LMYGLPDSSH VGKHEYFMHA TDKGGLSAVD
AFEIHVHRRP QGDRAPARFK AKFVGDPALV LNDIHKKIAL VKKLAFAFGD RNCSTITLQN ITrg
```

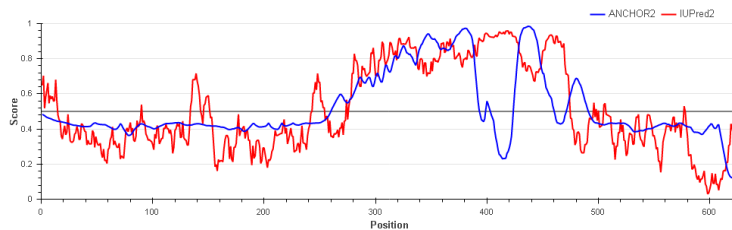# Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

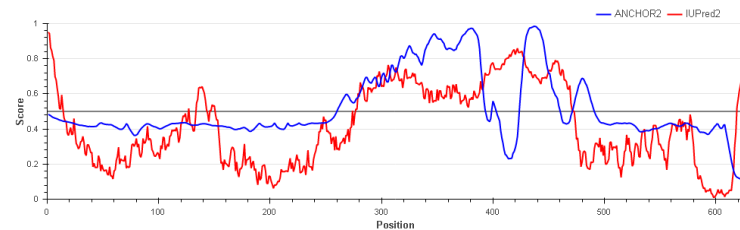**Analisi bioinformatica: A)   disorder prediction:**

3)   https://iupred2a.elte.hu/

stima l'energia totale di interazione tra singole coppie di residui,
(IPOTESI: sequenze disordinate non formano sufficienti interazioni stabilizzanti fra residui)
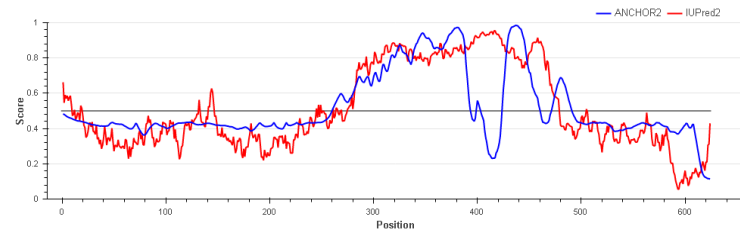
long disorder

short disorder

structured domains

```
Number of globular domains: 2
        globular domain   1.   21-257
        globular domain   2.   480-623
hwpsepseav rdwenqleas MHSVLSDLHE AVPTVVGIPD GTAVVGRSFR VTIPTDLIAS
SGDIIKVSAA GKEALPSWLH WDSQSHTLEG LPLDTDKGVH YISVSATRLG ANGSHIPQTS
SVFSIEVYPE DHSELQSVRT ASPDPGEVVS SACAADEPVT VLTVILDADL TKMTPKQRID
LLHRMRSFSE VELHNMKLVP VVNNRLFDMS AFMAGPGNAK KVVENGALLS WKLGCSLNQN
SVPDIHGVEA PAREGAMsaq lgypvvgwhi ankkpplpkr vrrqihatpt pvtaigpptt
aiqeppsriv ptptspaiap ptetmappvr dpvpgkptvt irtrgaiiqt ptlgpiqptr
vseagttvpg qirptmtipg yveptavatp pttttkkprv stpkpatpst dstttttrrp
tkkprtprpv prvttkvsit rletaspptr irtttsgvpr ggepnqrpel knhidrvdaW
VGTYFEVKIP SDTFYDHEDT TTDKLKLTLK LREQQLVGEK SWVQFNSNSQ LMYGLPDSSH
VGKHEYFMHA TDKGGLSAVD AFEIHVHRRP QGDRAPARFK AKFVGDPALV LNDIHKKIAL
VKKLAFAFGD RNCSTITLQN ITRg
```

Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Analisi bioinformatica: B) Predizione della struttura secondaria

- **http://bioinf.cs.ucl.ac.uk/psipred/**

- **http://www.compbio.dundee.ac.uk/jpred/index_up.html**
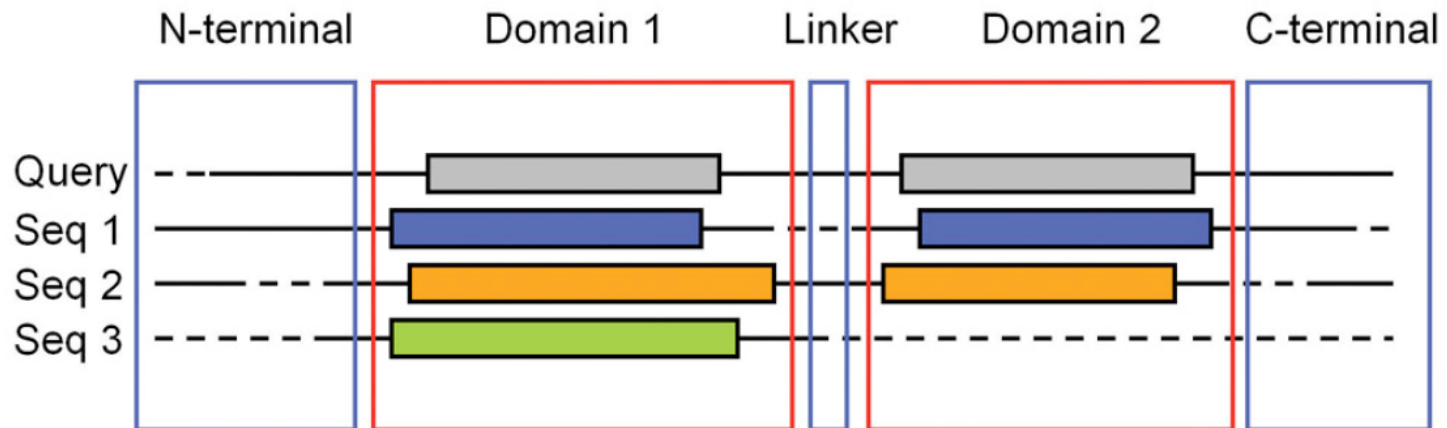
Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Analisi bioinformatica: C)   Allineamento con omologhi

1)   basato su struttura I

MSA:
- **https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome**
- **https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml (Conserved Domains Database)**

MSA+p-HMM (Hidden Markov Model):
- **https://pfam.xfam.org/**
- **http://smart.embl-heidelberg.de/ (Simple Modular Architecture Research Tool)**
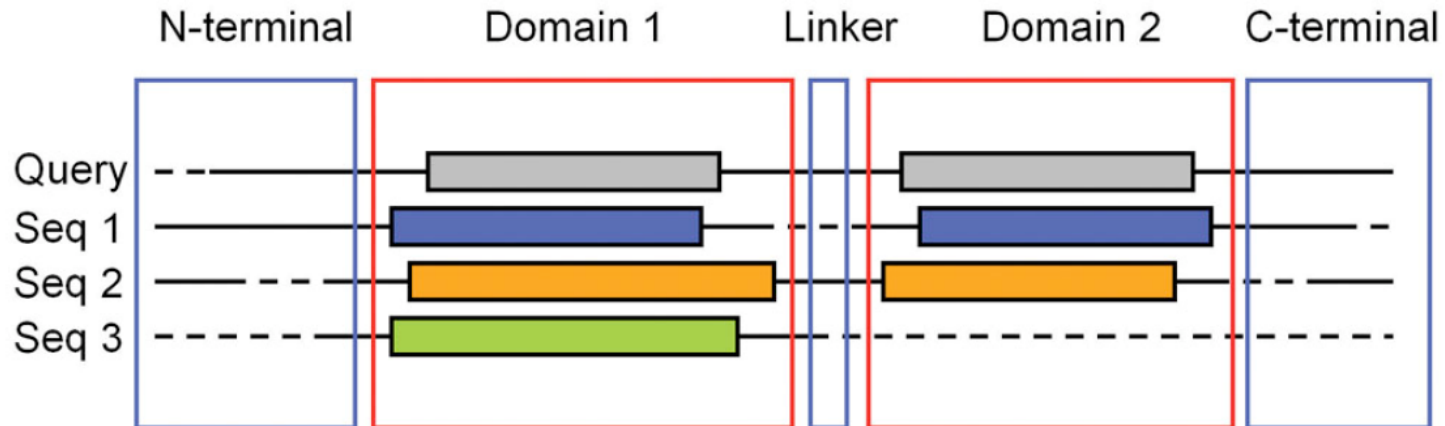
Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Analisi bioinformatica: C)   Allineamento con omologhi

2)   basato su struttura III

- **https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins (Database PDB)**

- **http://www.ebisu.co.uk/chemogenomix.com/chemogenomix/GenThreader.html**

Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Analisi bioinformatica: C)   Allineamento con omologhi

3)   basato su approcci ibridi

https://www.ics.uci.edu/~baldig/dompro.html (**Protein Domain Prediction)**
Homology (PSSM), Secondary Structure, Solvent Accessibility

http://bioinf.cs.ucl.ac.uk/web_servers/psipred_server/dompred_overview/
(Domain Prediction)
PSSM, Fold Recognition (Secondary Structure)

DomainDiscovery:
Position Specific Scoring Matrix (PSSM), Secondary Structure, Solvent Accessibility,
Domain Linker Index

Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Metodi sperimentali: D)   Definizione dei domain boundaries

1.   **Limited Proteolysis Mass Spectrometry (LP-MS)**
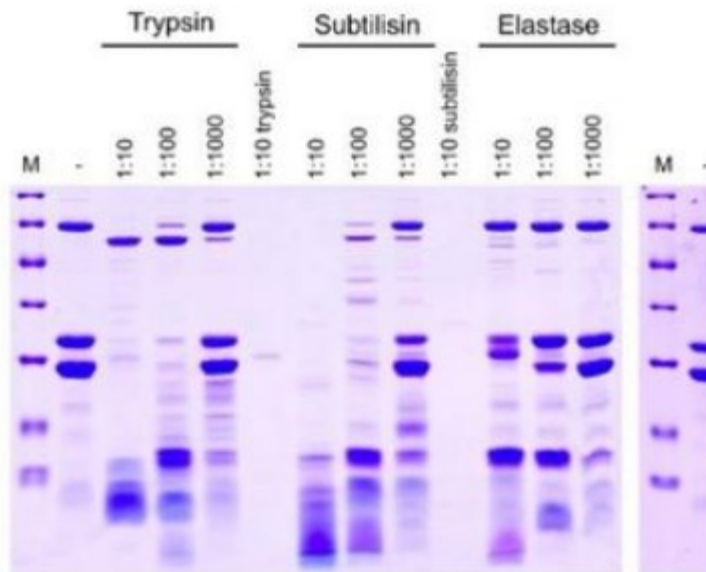
Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Metodi sperimentali: D)   **Definizione dei domain boundaries**

1. **Limited Proteolysis Mass Spectrometry (LP-MS)**

Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Metodi sperimentali: D)   Definizione dei domain boundaries

2)   Hydrogen/Deuterium Exchange (DX-MS)
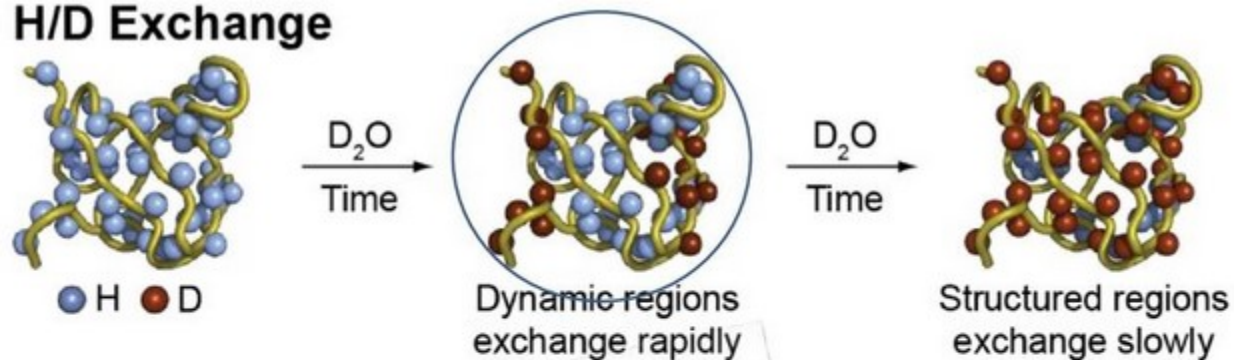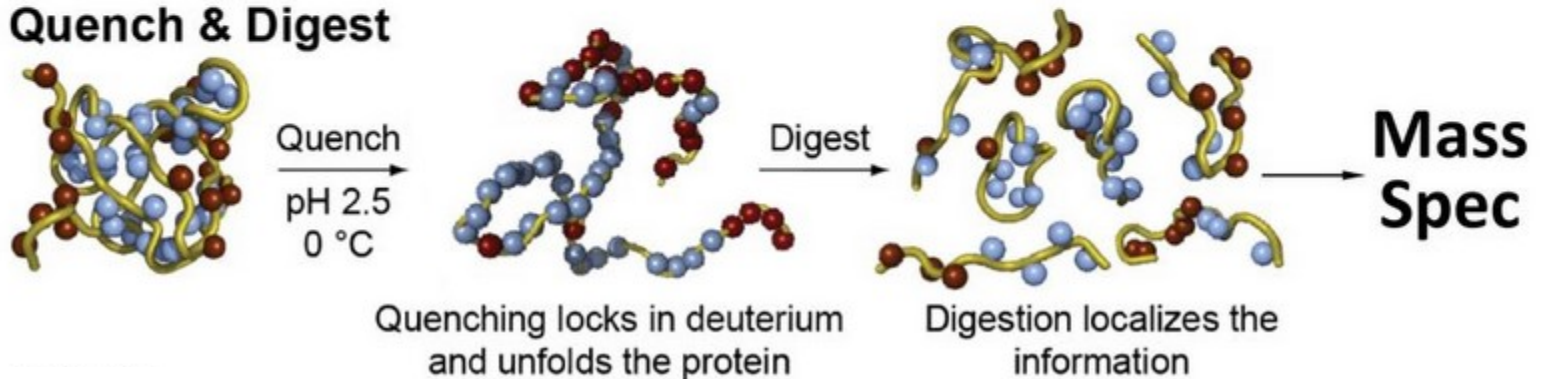
Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Metodi sperimentali: D)   Definizione dei domain boundaries

3)   Protein Painting
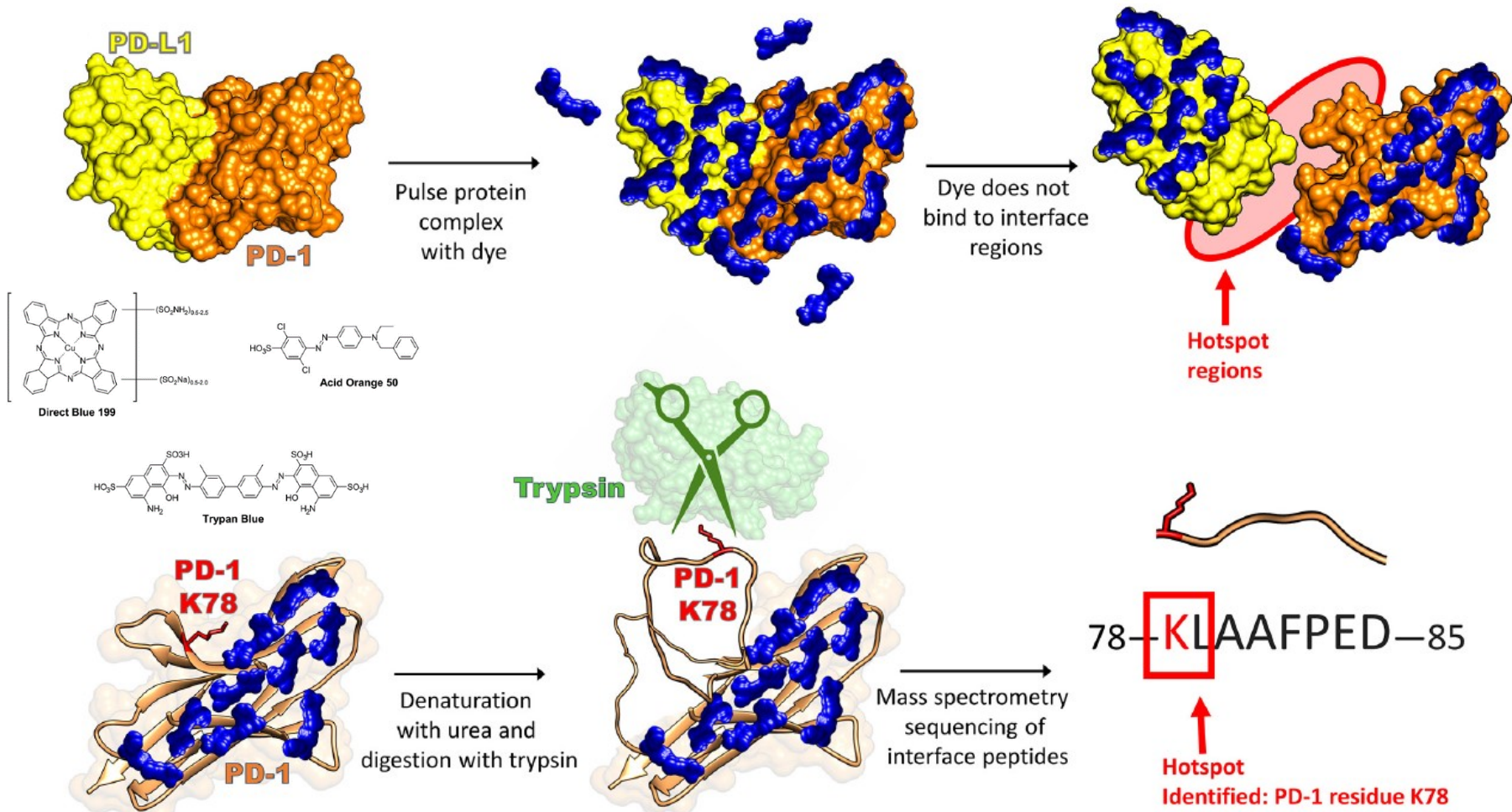
Espressione di proteine ricombinanti per studi strutturali
Design del costrutto

Espressione di singoli domini: CAVEAT

Domain swapping

Stefin B swapped dimer

MNEI/Stefin B
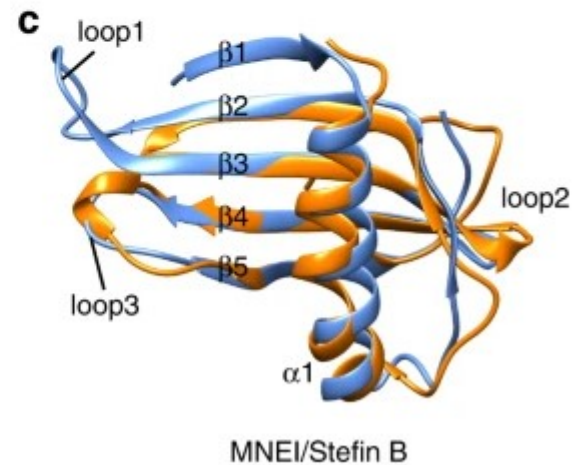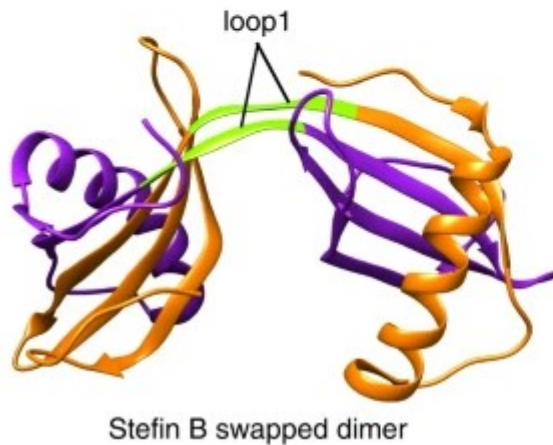
Domain swapped dimer: **Folding non nativo**

Espressione di proteine ricombinanti per studi strutturali
Scelta del sistema ospite

Espressione di proteine ricombinanti per studi strutturali
Scelta della cellula ospite:
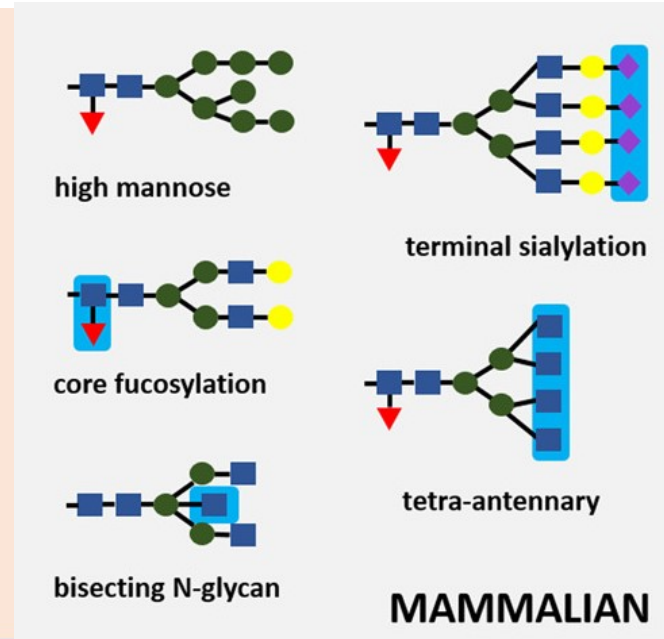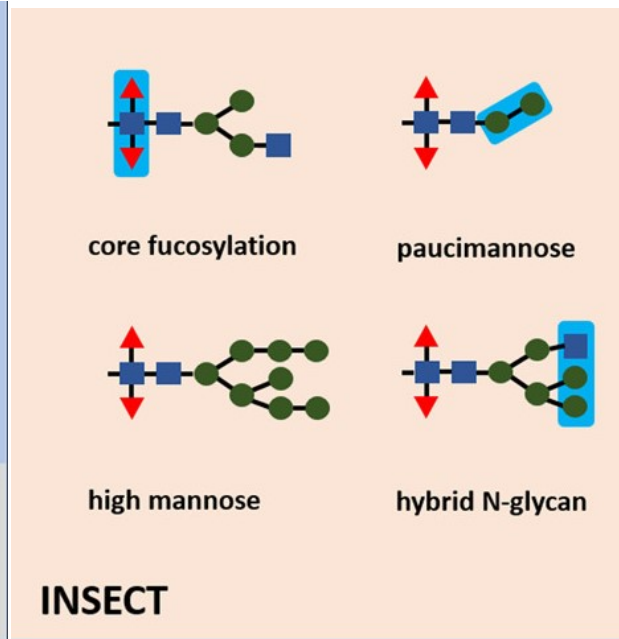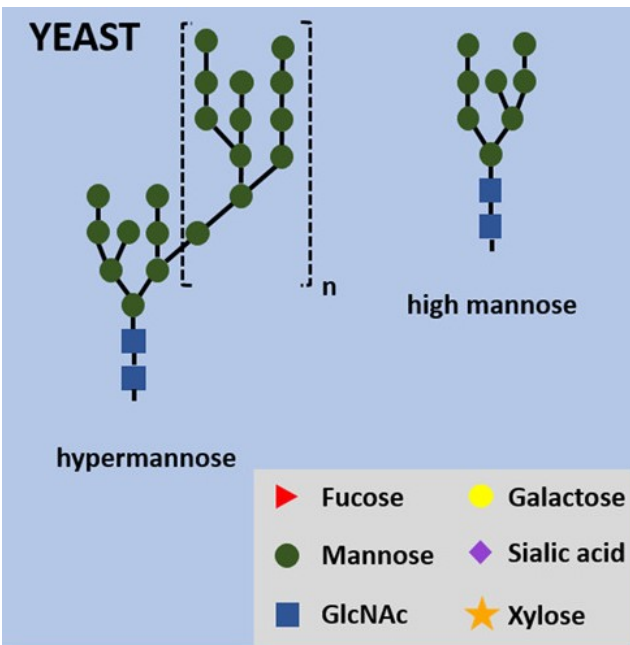
Modifiche post traduzionali

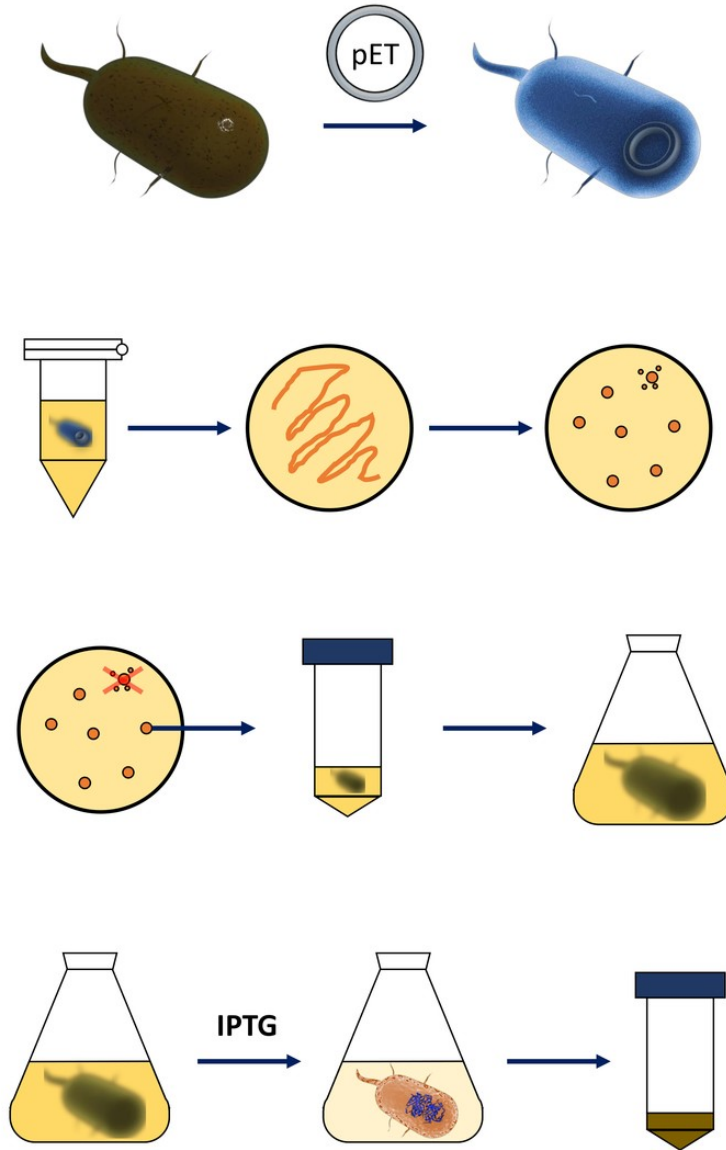# Espressione di proteine ricombinanti per studi strutturali
Scelta della cellula ospite:

Modifiche post traduzionali

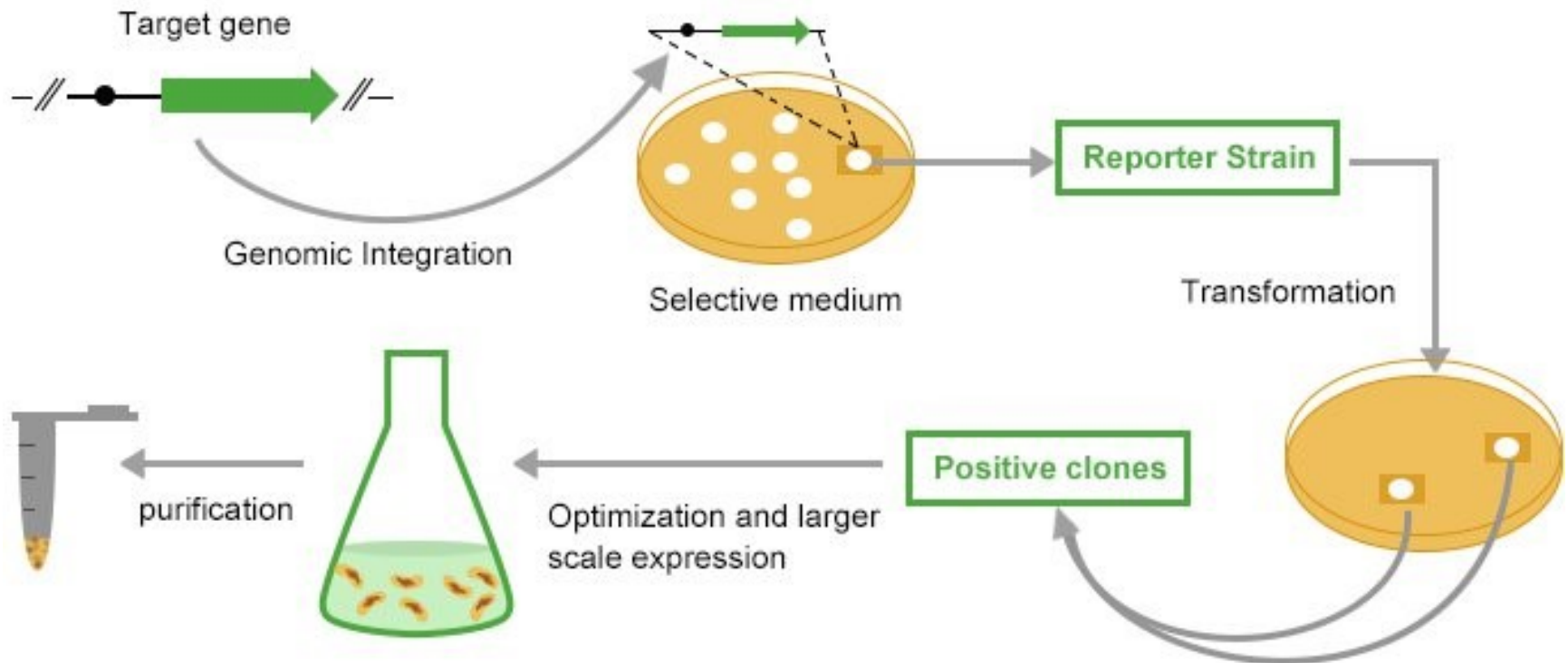# Espressione di proteine ricombinanti per studi strutturali
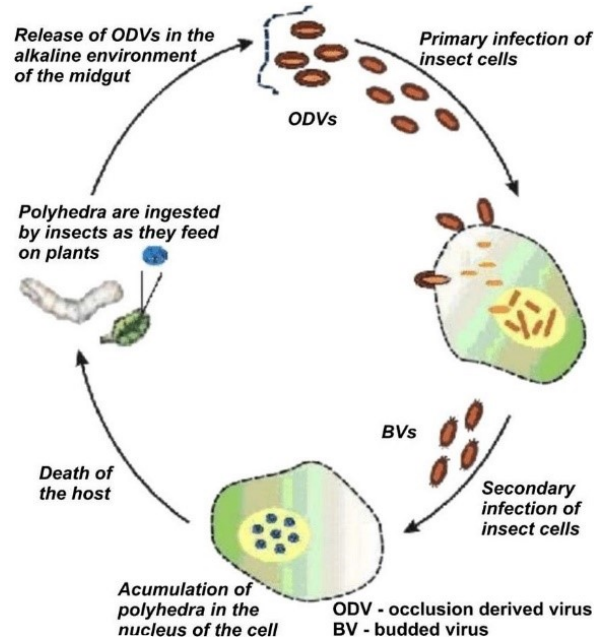## Scelta del sistema ospite: Procarioti

Espressione di proteine ricombinanti per studi strutturali
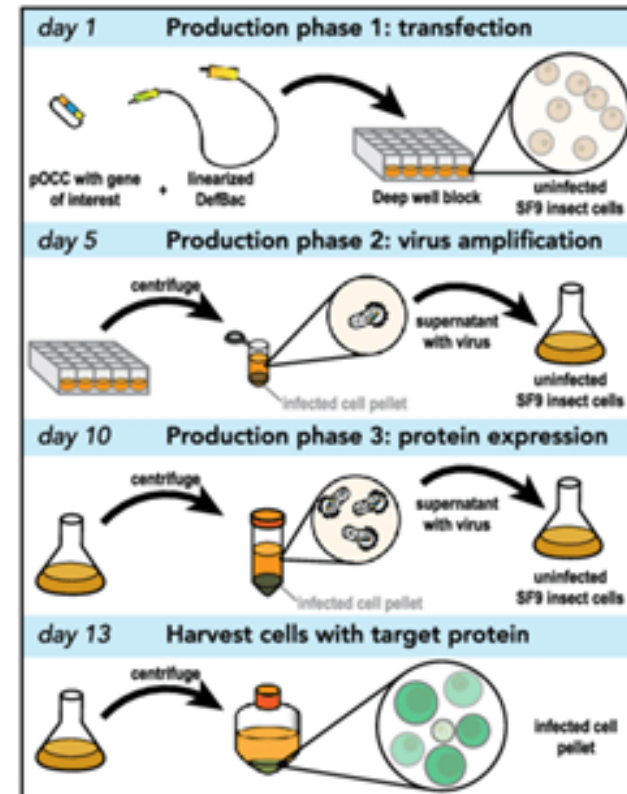## Scelta del sistema ospite: Lievito

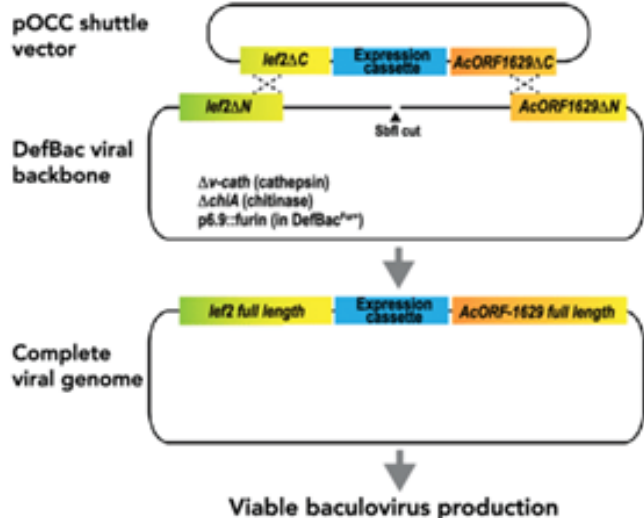# Espressione di proteine ricombinanti per studi strutturali
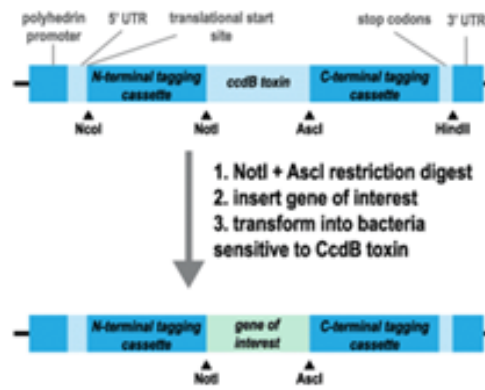## Scelta del sistema ospite: Baculovirus
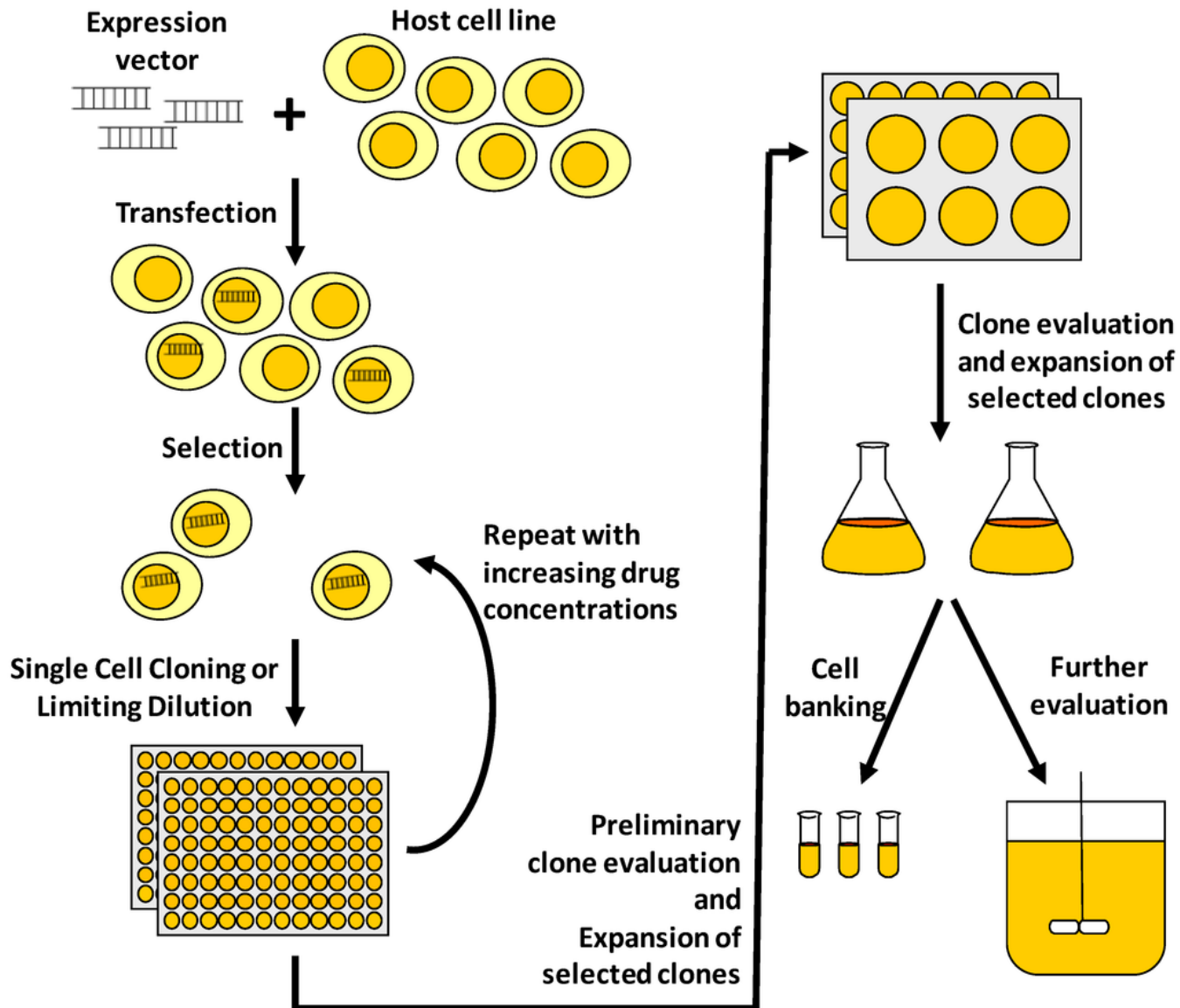
Espressione di proteine ricombinanti per studi strutturali
Scelta del sistema ospite: mammalian cells

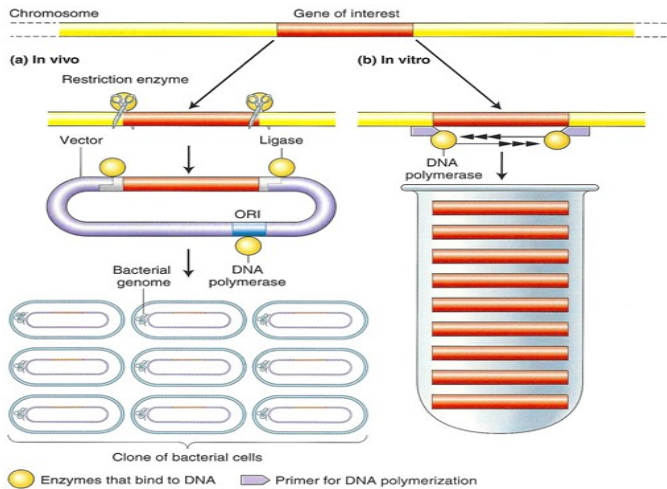# Espressione di proteine ricombinanti per studi strutturali
Scelta del sistema ospite

| Expression system	Most common application	Advantages	Challenges
Mammalian	• Functional assays • Structural analysis • Antibody production • Expression of complex proteins • Protein interactions • Virus production	• Highest-level protein processing • Can produce proteins either transiently, or by stable expression • Robust optimized transient systems for rapid, ultrahigh-yield protein production	• Gram-per-liter yields only possible in suspension cultures • More demanding culture conditions
Insect	• Functional assays • Structural analysis • Expression of intracellular proteins • Expression of protein complexes • Virus production	• Similar to mammalian protein processing • Can be used in static or suspension culture	• More demanding culture conditions than prokaryotic systems • Production of recombinant baculovirus vectors is time consuming
Yeast	• Structural analysis • Antibody generation • Functional analysis • Protein interactions	• Eukaryotic protein processing • Scalable up to fermentation (grams per liter) • Simple media requirements	• Fermentation required for very high yields • Growth conditions may require optimization
Bacterial	• Structural analysis • Antibody generation • Functional assays • Protein interactions	• Scalable • Low cost • Simple culture conditions	• Protein solubility • May require protein-specific optimization • May be difficult to express some mammalian proteins
Algal	• Studying photosynthesis, plant biology, lipid metabolism • Genetic engineering • Biofuel production	• Genetic modification and expression systems for photosynthetic microalgae • Superb experimental control for biofuel, nutraceuticals, and specialty chemical production • Optimized system for robust selection and expression	• Nascent technology • Less developed compared to other host platforms
Cell-free	• Toxic proteins • Incorporation of unnatural label or amino acids • Functional assays • Protein interactions • Translational inhibitor screening	• Open system; able to add unnatural components • Fast expression • Simple format	• Scaling above multimilligram quantities may not be costly

Espressione di proteine ricombinanti per studi strutturali
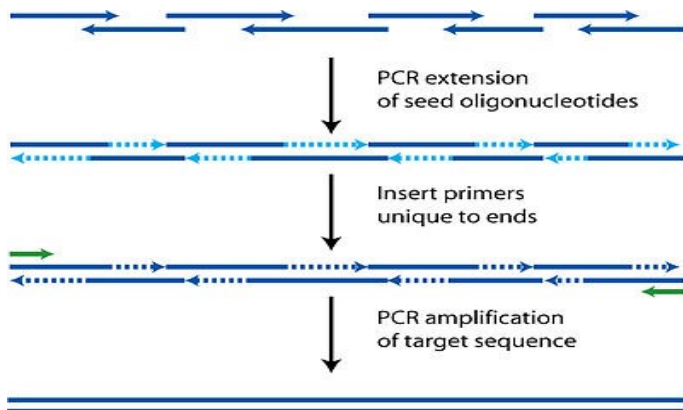## cDNA isolation:

genomic fishing

fishing from mRNA
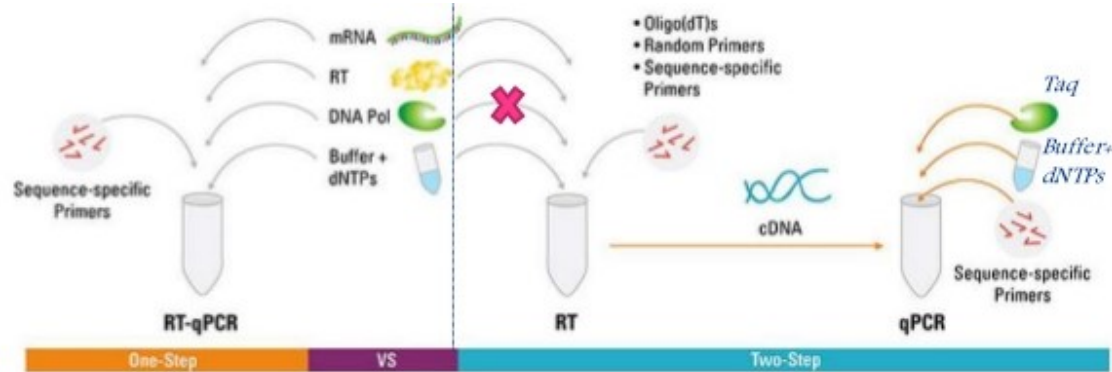
strategia ibrida sintesi/PCR
(overlap assembly PCR)
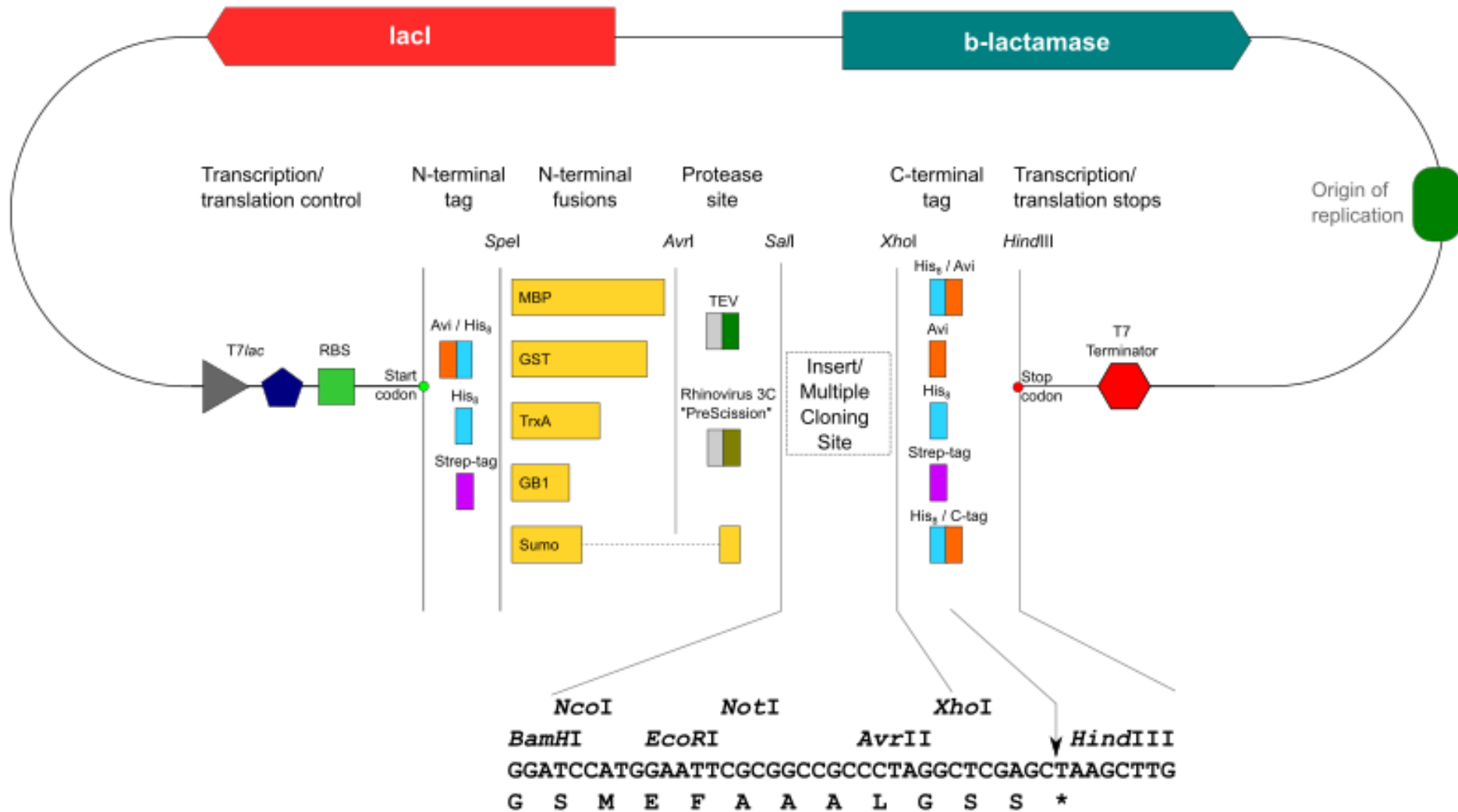
sintesi genica

Espressione di proteine ricombinanti per studi strutturali
Scelta del vettore :

elementi principali

Espressione di proteine ricombinanti per studi strutturali
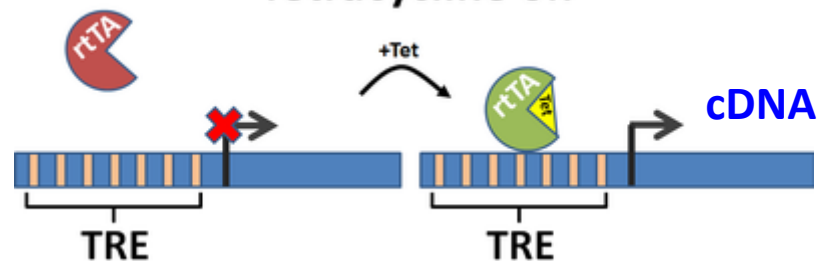## Scelta del vettore : PROMOTORE

In base a

- Ospite
- Forza
- Leakiness
- Lag time
- Costitutivo/inducibile

Positive inducible

activator

Inducer addition
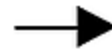
Transcription activated
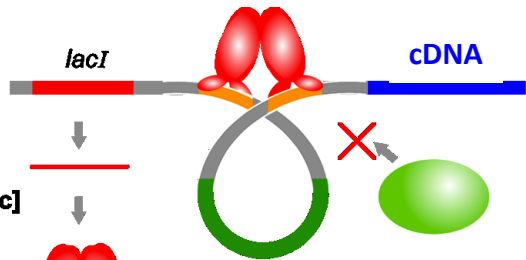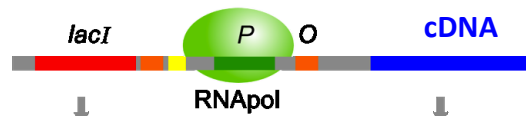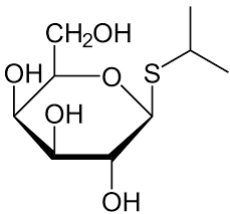
Tetracycline on

+Tet

rtTA

cDNA

TRE

TRE

Espressione di proteine ricombinanti per studi strutturali
Scelta del vettore : PROMOTORE

Espressione di proteine ricombinanti per studi strutturali
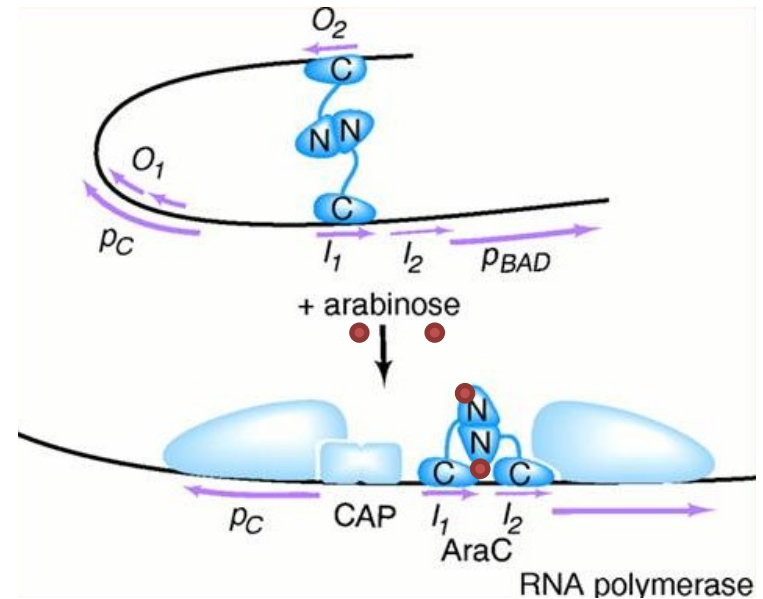## promotori batterici

Promoter	Description	Expression	Additional considerations
T7	Promoter from T7 bacteriophage	Constitutive, but requires T7 RNA polymerase.	
T7lac	Promoter from T7 bacteriophage plus lac operators	Negligible basal expression. Requires T7 RNA polymerase, which is also controlled by lac operator.	Very tightly regulated by the lac operators, tunable.
Sp6	Promoter from Sp6 bacteriophage	Constitutive, but requires SP6 RNA polymerase.	SP6 polymerase has a high processivity.
araBAD	Promoter of the arabinose metabolic operon	Inducible by arabinose repression in the presence of glucose or by competitive binding of the anti-inducer fucose	Weaker. Low basal expression; not tunable
trp	Promoter from E. coli tryptophan operon	Repressible with high levels of cellular trp.	
lac	Promoter from lac operon	Constitutive in the absense of lac repressor (lacI or lacIq). Can be induced by IPTG or lactose.	Leaky promoter weak expression tunable
Ptac	Hybrid promoter of lac and trp	Regulated like the lac promoter	Contains -35 region from trpB and -10 region from lac. Very tight regulation. tunable
pL	Promoter from bacteriophage lambda	temperature regulable	Often paired with the temperature sensitive cI857 repressor.
T3	Promoter from T3 bacteriophage	Constitutive requires T3 RNA polymerase	

Espressione di proteine ricombinanti per studi strutturali
sistemi batterici

DE3 system
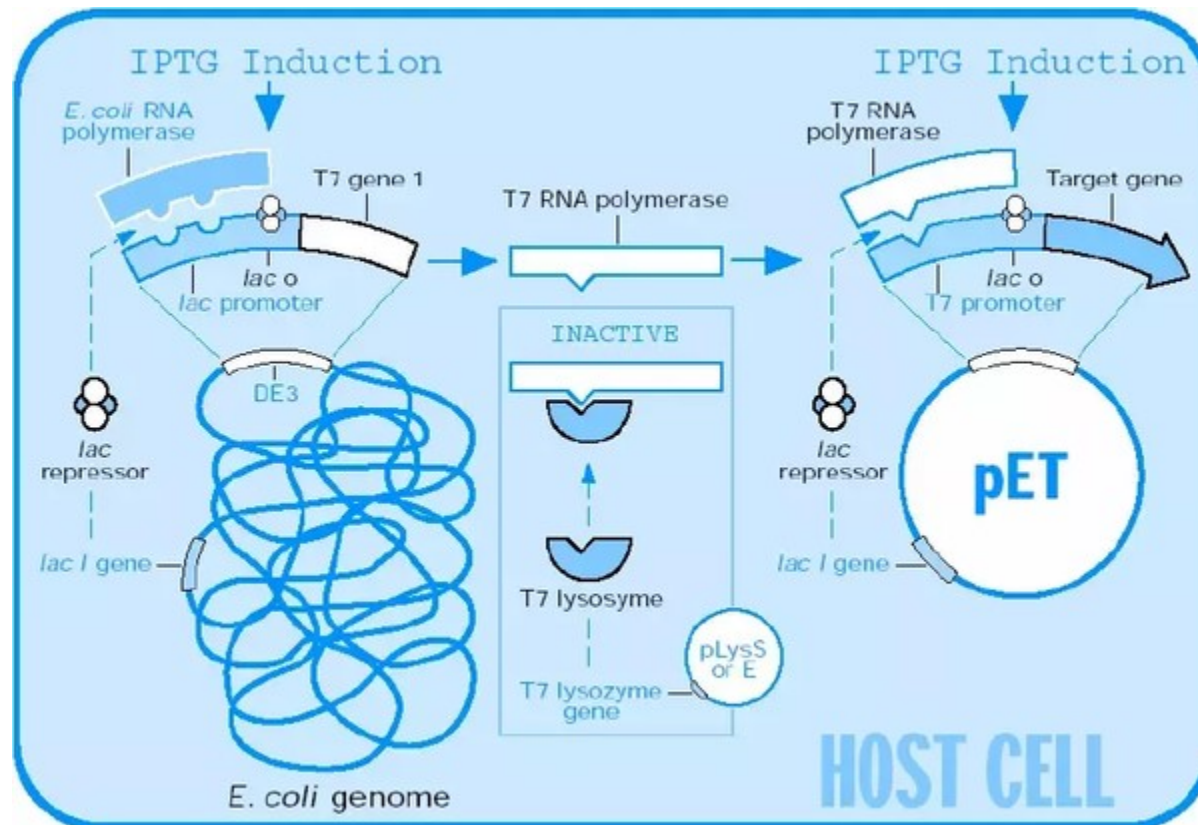
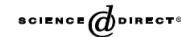**Espressione inducibile**
Autoinduzione

Espressione di proteine ricombinanti per studi strutturali
Autoinduzione

In alcuni media (mix glucosio e lattosio)
è possibile indurre l'espressione quando
i batteri raggiungono la saturazione.
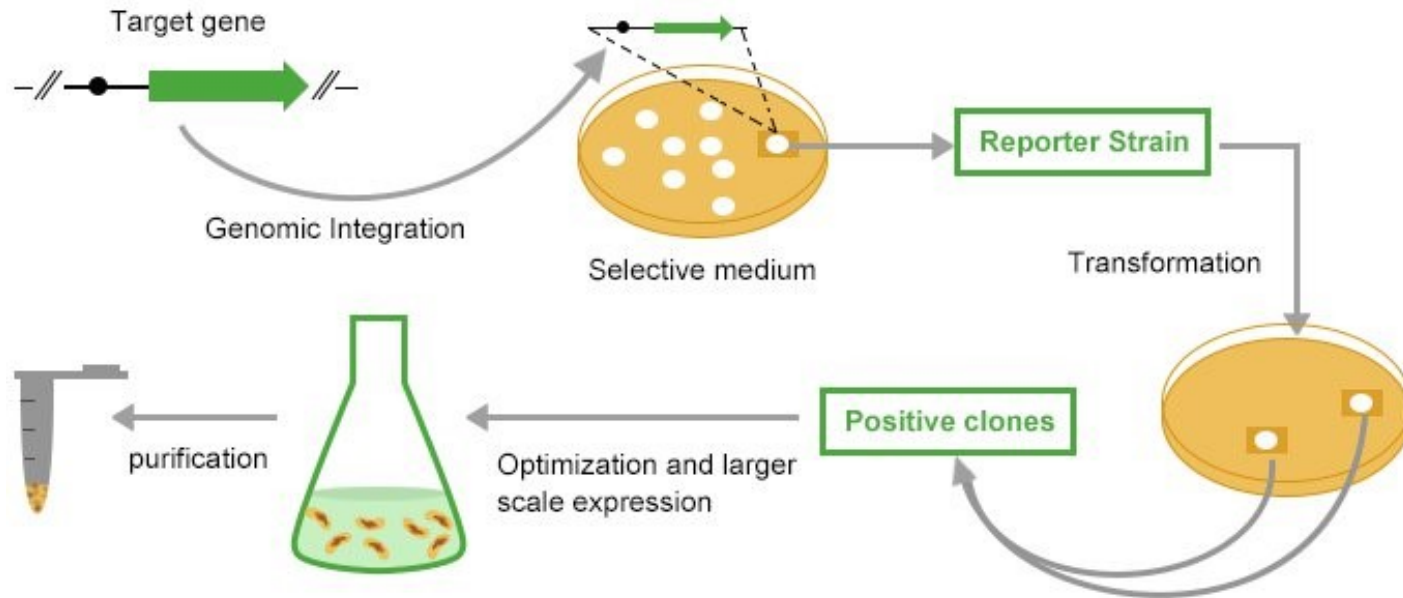
Protein production by auto-induction in high-density shaking cultures
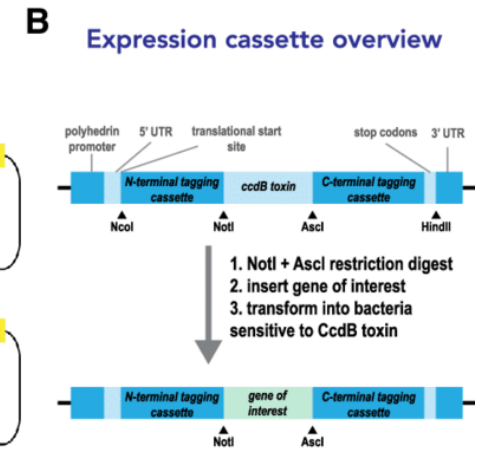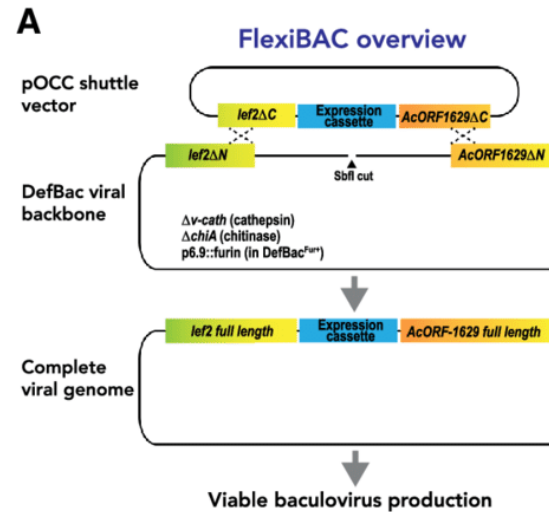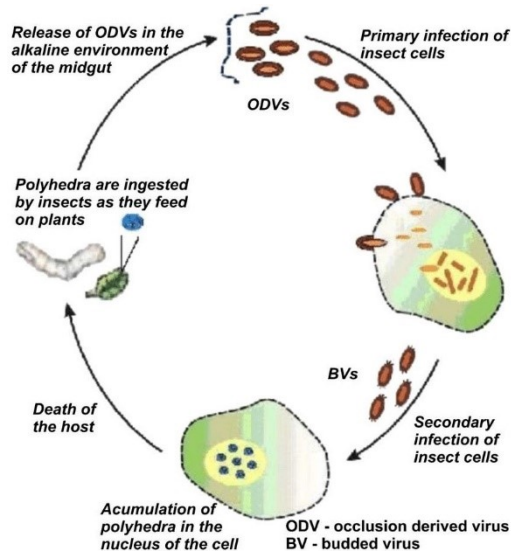
F. William Studier *
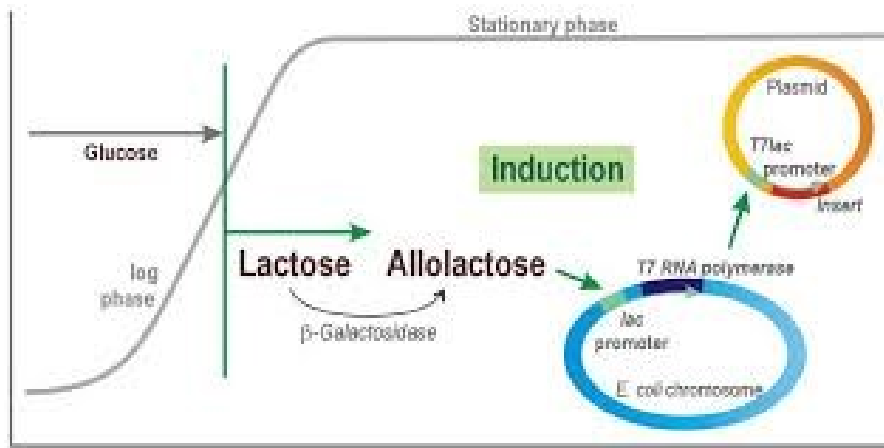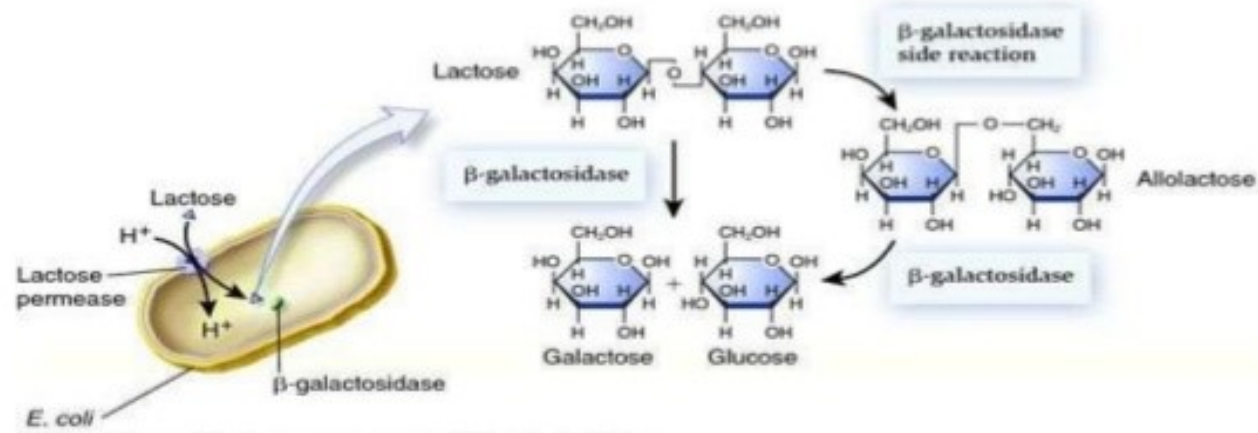
Si possono raggiungere da 2 a 3X OD600nm

Espressione di proteine ricombinanti per studi strutturali
promotori di Lievito

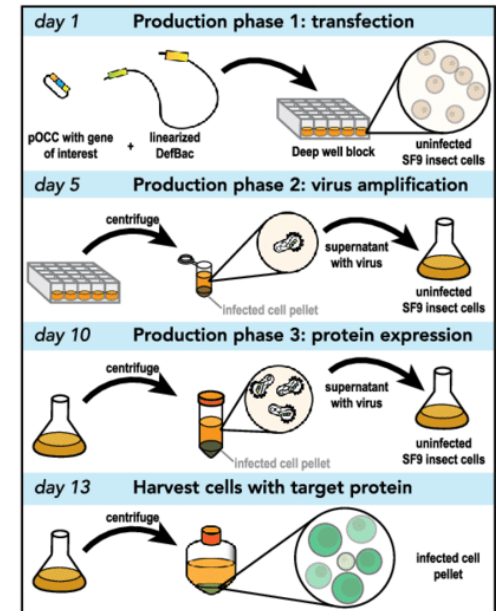| Promoter	Description	Expression	Additional considerations
GAL1, 10	adjacent, divergently transcribed promoters	Inducible with galactose; repressible with glucose	Used independently or together. Regulated by GAL4, 80.
TEF1	from transcription elongation factor	Constitutive	Analogous to mammalian EF1a promoter.
GDS	Strong from glyceraldehyde 3-phosphate dehydrogenase	Constitutive	Very strong
ADH1	From alcohol dehydrogenase I	Repressed by ethanol	Full length version is strong. Truncated constitutive with lower expression.

Release of ODVs in the alkaline environment of the midgut

Primary infection of insect cells

ODVs

Polyhedra are ingested by insects as they feed on plants

BVs

Death of the host

Secondary infection of insect cells

Acumulation of polyhedra in the nucleus of the cell

ODV - occlusion derived virus
BV - budded virus

A **FlexiBAC overview**

pOCC shuttle vector

lef2ΔC — Expression cassette — AcORF1629ΔC

lef2ΔN — AcORF1629ΔN

SbfI cut

DefBac viral backbone

Δv-cath (cathepsin)
ΔchiA (chitinase)
p6.9::furin (in DefBac^Furin)

Complete viral genome

lef2 full length — Expression cassette — AcORF-1629 full length

Viable baculovirus production

B **Expression cassette overview**

polyhedrin promoter | 5' UTR | translational start site | stop codons | 3' UTR

N-terminal tagging cassette | ccdB toxin | C-terminal tagging cassette

NcoI | NotI | AscI | HindII

1. NotI + AscI restriction digest
2. insert gene of interest
3. transform into bacteria sensitive to CcdB toxin

N-terminal tagging cassette | gene of interest | C-terminal tagging cassette

NotI | AscI

Baculovirus Genome	Deletion of *cathepsin*
	Deletion of *chitinase*
	Double deletion of *cathepsin* and *chitinase*
	Triple deletion of *p26,p10* and *p74*
	BmNPV *fp25k* replaced by AcMNPV *fp25k*
Insect Promoters	*polh* promoter
	p10 promoter
	p6.9 promoter
	vp39 promoter
	hybrid *vp39* and *polh* promoter
	polh promoter enhanced by Homologous Region 1
	vp39 promoter enhanced by Homologous Region 3
Signal Peptides	honeybee melittin signal peptide
	cecropin B signal peptide
	Staphylococcal protein A signal peptide
	human azurocidin signal peptide
	B.mori bombyxin(*bx*) signal peptide
	B.mori prophenoloxidase-activating enzyme(*ppae*) signal peptide

C **Tagging cassette options**

Fluorophores
monoGFP	eGFP
mCherry	mCitrine
tagRFP	mKate2
mRuby3	

Affinity tags
6xHis	10xHis
MBP	pfuMBP
GST	S-tag
TwinStrep	

Protease cleavage sites
3C	TEV

Other
SNAP	SUMO3
Cry2	Sortase tag
HBM	Avi Tag
YBBR	GP64

D **Protein expression timeline**

day 1 **Production phase 1: transfection**

pOCC with gene of interest + linearized DefBac → Deep well block → uninfected SF9 insect cells

day 5 **Production phase 2: virus amplification**

centrifuge → infected cell pellet → supernatant with virus → uninfected SF9 insect cells

day 10 **Production phase 3: protein expression**

centrifuge → infected cell pellet → supernatant with virus → uninfected SF9 insect cells

day 13 **Harvest cells with target protein**

centrifuge → infected cell pellet

Espressione di proteine ricombinanti per studi strutturali
promotori di mammalian cells

Promoter	origin	Expression	comments
CMV	Strong from the human cytomegalovirus	Constitutive	May contain an enhancer. Silenced in some cell types.
EF1a	Strong from human elongation factor 1α	Constitutive	High expression
SV40	from the simian vacuolating virus 40	Constitutive	May include an enhancer.
PGK1 (human or mouse)	from phosphoglycerate kinase gene.	Constitutive	vary by cell type. Resist downregulation due to methylation or deacetylation.
Ubc	from the human ubiquitin C gene	Constitutive	Ubiquitous
human beta actin	from beta actin gene	Constitutive	Ubiquitous
TRE	Tetracycline response element promoter	Inducible	minimal promoter with low basal activity and several tetracycline operators.
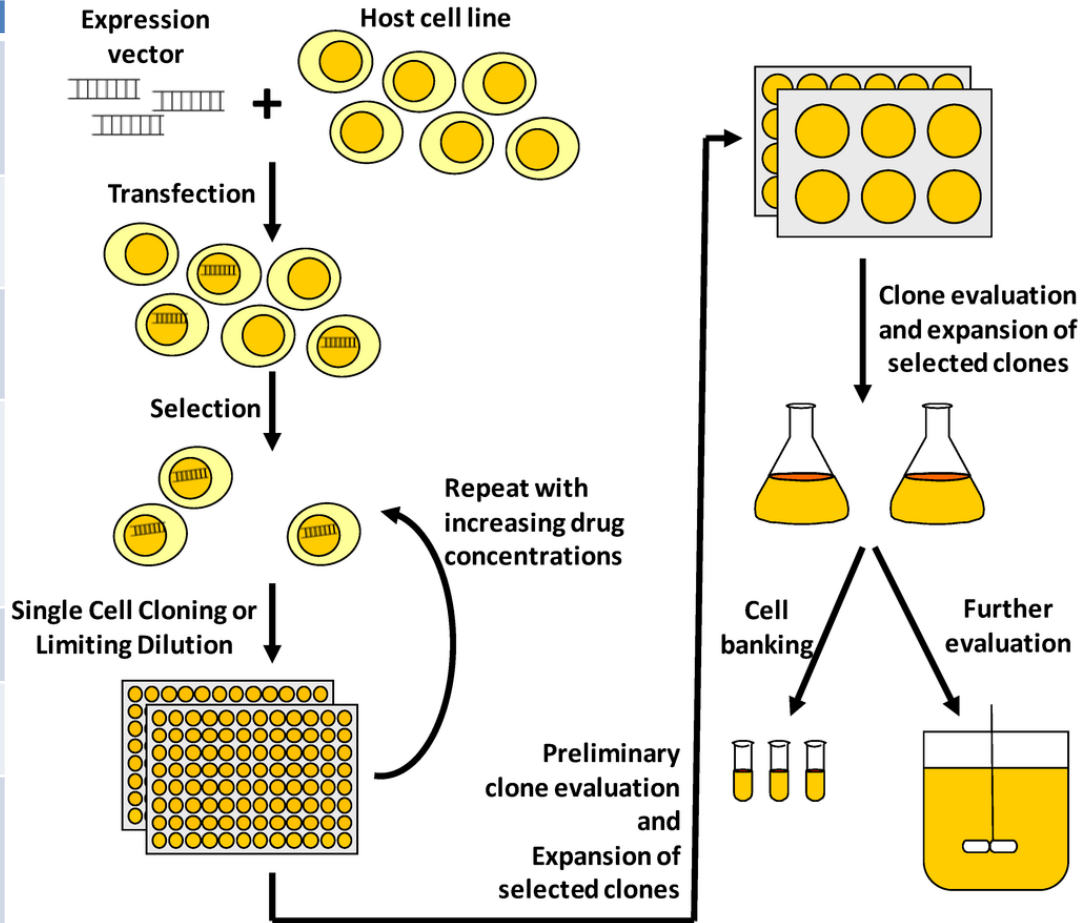
Espressione di proteine ricombinanti per studi strutturali
Scelta del vettore: TAGs&clevage sites

Espressione di proteine ricombinanti per studi strutturali
Scelta del vettore: TAGs

Tag	Protein source	Residue size (aa/ ~kDa)	Tag placement	Source organism	Ligand/Matrix	Purification condition	Suitable expression system	Advantages	Limitations
Solubility enhancing tags									
MBP	Maltose binding protein	396/43	N- or C-terminus	E. coli	Cross linked amylose	10 mM Maltose	E. coli	Enhanced solubility of target, do not leave extraneous amino acids	Can decrease expression yield
GST	Glutathione-S-transferase	211/26	N- or C-terminus	Schistosoma japonicum	Glutathione	10-20 mM reduced Glutathione	Bacteria, Yeast, insect, mammalian cells	Enhanced solubility of target, Highly specific binding	Can leach from column
SUMO	Small ubiquitin modifier	100/18	N-terminus	H. sapiens	NA	N/A	Bacteria, Yeast, Insect, Mammalian cells	Enhanced target solubility, unique tag removal method	Do not aid in purification
TrxA	Thioredoxin A	109/12	N- or C-terminus	E. coli	Phenylarsinine oxide	5-1000 mM β-mercapto-ethanol	E. coli	Helps in crystallization, Aid in refolding of target protein	Do not aid in purification
NusA	N-utilization substance A	495/55	N- or C-terminus	E. coli	N/A	N/A	Bacteria	Enhanced solubility	Large size decreases yield
Mistic	Mistic tag	110/13	N- or C-terminus	Bacillus subtilis	NA	N/A	Bacteria	Commendable tag for membrane proteins	Do not aid in purification and crystallization
HaloTag7	Haloalkane dehalogenase	~300/34	N- or C-terminus	Rhodococcus/ engineered	Chloroalkane coupled agarose HaloLink™ resin	In-column proteolytic digestion	Insect cells, Mammalian cells	Increased expression, solubility, purification yield and high purity, in-vivo imaging and multiplexing	Application dependent ligand requirement; Functional inactivation of target protein
Fh8	Ca²⁺ binding recombinant protein	69/8.0	N-terminus	Faciola hepatica	Phenyl sepharose	N/A	E. coli	Enhanced expression & solubility, aid in purification on HIC	
DsbA	Protein disulphide isomerase A	209/~21	N-terminus	E. coli	NA	N/A	E. coli	Enhanced solubility, native disulfide bond formation in target	
Skp	Small E. coli chaperone	162/17.9	N-terminus	E. coli	N/A	N/A	E. coli	Enhanced solubility	Do not aid in purification
GB1	IgG domain B1 of Streptococcus Protein G	56/7.5	N-terminus	Streptococcus aureus	N/A	N/A	Bacteria	Enhanced refolding	
ZZ domain	IgG binding ZZ domain of protein A	/31	N- or C-terminus	Staphylococcus aureus	N/A	N/A	Bacteria, Yeast	Enhanced refolding	
Stimulus responsive tags									
BRT17	β-roll repeat of GGAGNDTLY	153/14.7	N-terminus	Synthetic	N/A	Precipitation in presence of 25 −75 mM Ca⁺⁺	Bacteria	Selective purification	Refolding of precipitated target
ELP	Elastin-like peptides	550 (110 repeats)/ ~25	N- or C-terminus	tandem repeats of VPGXG sequence	Reverse phase transition	Precipitation by temperature change and/ or > 1.5 M NaCl	Bacteria	Cost effective than conventional purification	Refolding of precipitated target

Espressione di proteine ricombinanti per studi strutturali
Scelta del vettore: TAGs

Tag	Protein source	Residue size (aa/ ~kDa)	Tag placement	Source organism	Ligand/Matrix	Purification condition	Suitable expression system	Advantages	Limitations
					RNAse A) agarose	buffer, pH 2/3 M MgCl₂			
Calmodulin Binding Protein	Skeletal muscle myosin light chain kinase	26/2.5	N- or C terminus	H. sapiens	Calmodulin	EGTA and >1.5 M NaCl	Bacteria	High specificity, purification yield, native elution	Unsuitable for eukaryotic systems
T₇ Tag		11/1.2	N-/C- or internal	Bacteriophage T₇	Anti-T₇ epitope monoclonal antibody	Low pH	Bacteria	Increased expression	Inclusion body formation, refolding efficiency of purified target is compromised
Chitin binding Protein	Chitinase A	51/6.0	N- or C- terminus	Bacillus circulans	Chitin	β-mercaptoethanol or DTT	Bacteria	High purity	Used with auto-cleavable intein tag, target, Target loose first methionine after tag removal
Affinity epitope tags									
Rho1D4	Last 9 aa of C-terminus of Bovine rhodopsin	9/1.0	C terminus	Synthetic	Monoclonal anti-Rho1D4	Low pH	Bacteria, Yeast, Mammalian cell lines	High specificity, membrane protein purification, Immunolocalization	
c-Myc	Human c-Myc protooncogene	11/1.2	N-/C- or internal	H. sapiens	Anti-Myc antibody (Mycl-9E10) coupled on agarose	Low pH	Mammalian cell lines	Phenomenal purity	Expensive due to limited reuse
PDZ	Ina D	80-90/- 9-10	N-/C- or internal	H. sapiens	Immobilized C-terminal PDZ ligand sequence	DTT or β-mercaptoethanol	Insect, Mammalian	Specificity and high affinity	
HA	Human influenza virus hemagglutinin protein	9/1.0	N-/C- or internal	Human influenza virus	Monoclonal antibody immobilized on agarose	Low pH or HA peptide	Mammalian cell lines	Useful in detection, isolation and purification	Affects cellular trafficking and folding
FLAG®	Engineered	8/1.01	N- or C-terminus	Synthetic	Anti-FLAG antibody immobilized on agarose	2-5 mM EDTA	Bacteria, Yeast, Insect, Mammalian	Minimal impact on target	Expensive & limited reuse

Solubility decreasing (inclusion body targeting) tags

KSI	Ketosteroid isomerase	125/13	N-terminus	*Pseudomonas testosterone*	N/A	N/A	*E coli*	In column purification, refolding and sorting	Tag removal uses CNBr
TrpΔLE	*Trp L* operon	17/27	N- or C-terminus	*E coli*	N/A	N/A	*E coli*	High expression of toxic and membranes proteins	Limited Refolding efficiency
PagP	β- barrel outer membrane protein			*Pseudomonas aruginosa*	N/A	N/A	Bacteria	Easier inclusion body solubilisation of toxic target, commercialized	Refolding efficiency

Affinity Purification tags

Poly Arg	N/A	5-6/0.8	C terminus	Synthetic	Cation exchange resin	Alkaline NaCl	Bacteria	General cation exchange resin for purification	Cryptic cleavage, altered tertiary structure of target
Poly His	N/A	6-10/ 0.84 −1.4	N- or C terminus	Synthetic	Ni^{2+}-NTA agarose	20-250 mM Imidazole/low pH	Bacteria	Minimal interference with target protein, well defined protocol and availability	Co-purification of contaminants
Strep II Tag	Bacterial peptide	8/1.06	N- or C terminus	Bacteria	Engineered streptavidin (Strep-tactin)	2-25 mM Desthiobiotin	Bacteria, Yeast, Plants Mammalian Insect cells	Immunodetection, specificity and versatility	
S-Tag	Bovine pancreatic RNAse A	15/1.75	N-/C- or internal	Bovine	S-protein (21 −124 aa of	3 M GITC; 0.2 M Potassium citrate	Bacteria	Unlikely interference with protein folding	Compromised purification

Espressione di proteine ricombinanti per studi strutturali
Scelta del vettore: siti di taglio per Proteasi

Protease	Cleavage site	Location	Residual aa	pH range	Chaotropic sensitivity	Salt sensitivity	Enzyme/substrate ratio
TEV, AcTEV, ProTEV	ENLYFQˆ(G/S)	N	G/S	5.5	2 M urea	<0.1 M	1-3% (wt/wt)
		C	ExxYxQ	−8.5			
Thrombin (fIIa)	LVPRˆGS	N	G	5–10	<0.1 M urea	<0.1 M	1-10% (wt/wt)
		C	LVPR				
Factor Xa (fXa)	IEGRˆ x	N	None	6–9	<0.1 M urea	<0.15 M	1-10% (wt/wt)
	x ≠ R, P	C	IEGR				
Genenase I							
Enterokinase	DDDDKˆ x	N	None	7–8			0.1% (wt/wt)
		C	DDDDK				
SUMO proteases	C-terminus of SUMO ˆ xaa-	N only	None	6–9	<0.1 M GnCl	<0.5 M	0.1% (wt/wt)
	GG/yaa				2 M urea		
PreScission (human rhinovirus 3C protease)	LEVLFQˆ GP	N					
TVMV (Tobacco vein mottling virus protease)	ETVRFQGˆS	N	GP				
Immobilized subtilisin (BPN); Profinity eXact system	C-terminus of propeptide ˆ x EEDKLFAL	N only	None	7.2	2 M urea		1:1(mol:mol)
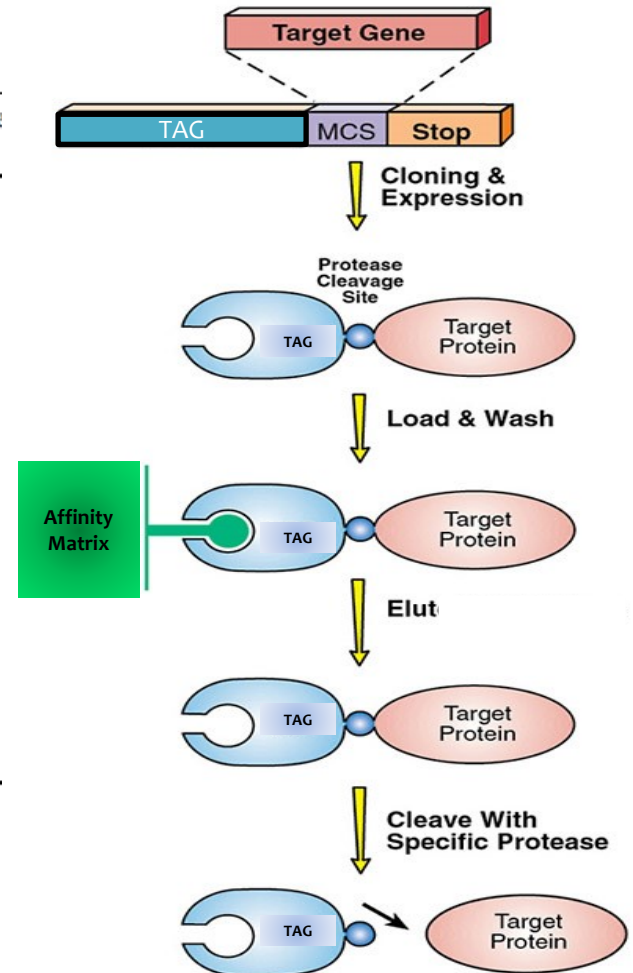
Espressione di proteine ricombinanti per studi strutturali
Clonaggio:

Ligation-dependent

Clonaggio per restrizione   **TA cloning**

MCS
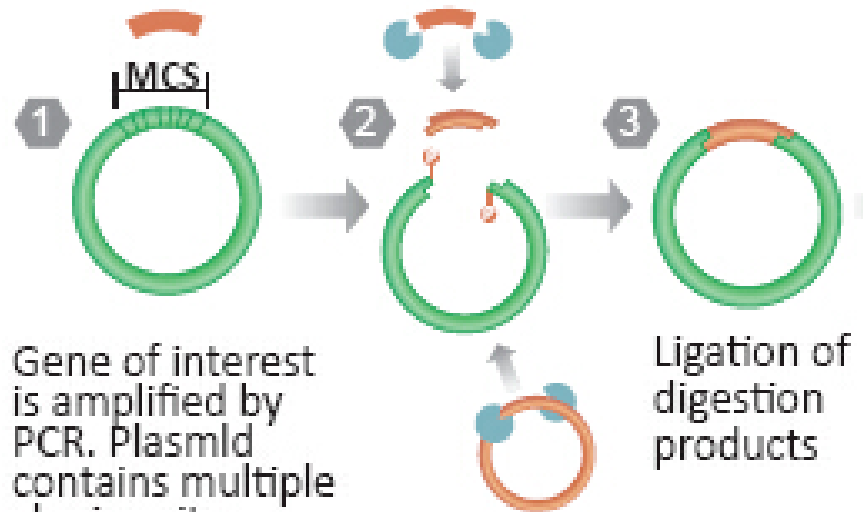
Gene of interest is amplified by PCR. Plasmid contains multiple cloning site, encompassing multiple restriction enzyme recognition sites.

Digestion of PCR product and vector by restriction enzymes. Dephosphorylation of vector may be required.

Ligation of digestion products
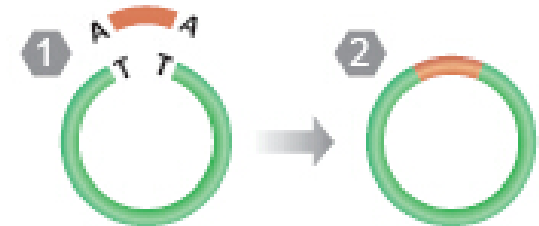
Gene of interest is amplified by PCR and contains adenosine overhangs. Plasmid contains thymine overhangs.

Ligation of gene and vector

Espressione di proteine ricombinanti per studi strutturali
Clonaggio:

Ligation-independent

RF-cloning
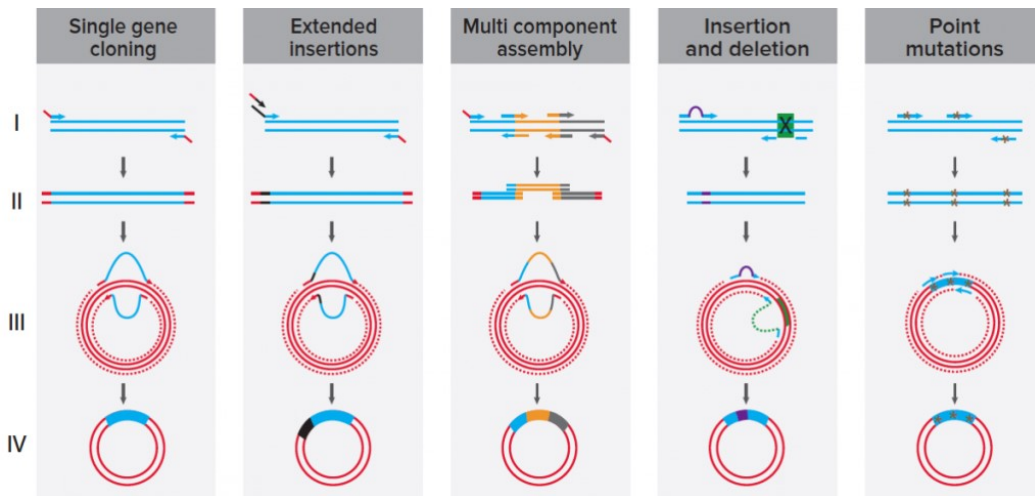
LI- cloning (LIC)
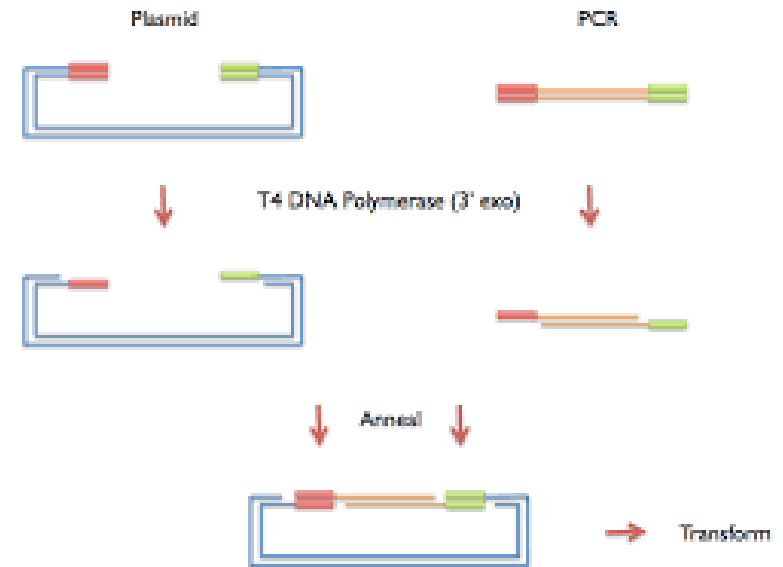
TOPO cloning:

Vaccinia Topoisomerasi I (InVitrogen)
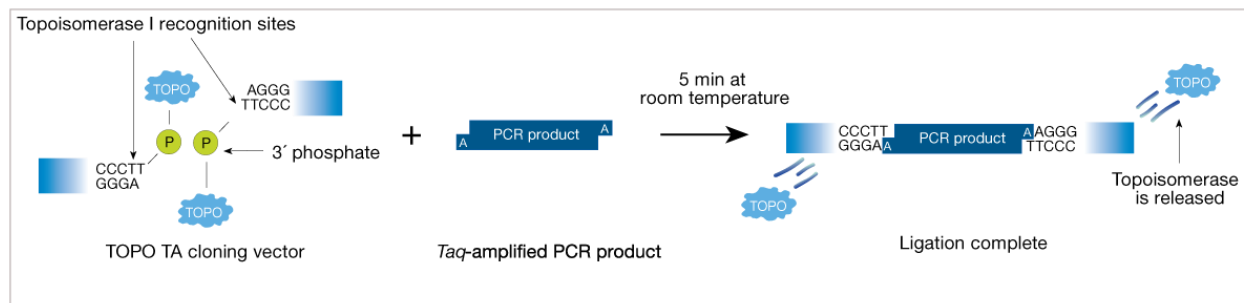
Espressione di proteine ricombinanti per studi strutturali
Clonaggio:

Ligation-independent

Gateway cloning:
fago λ − attB/P (Life technologies)
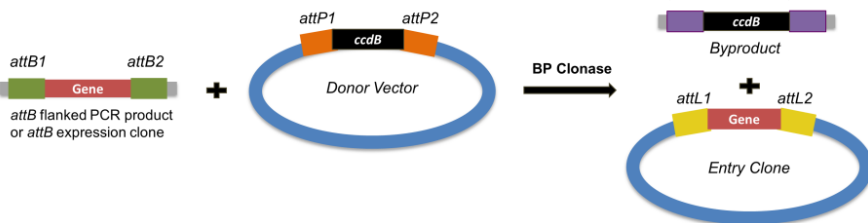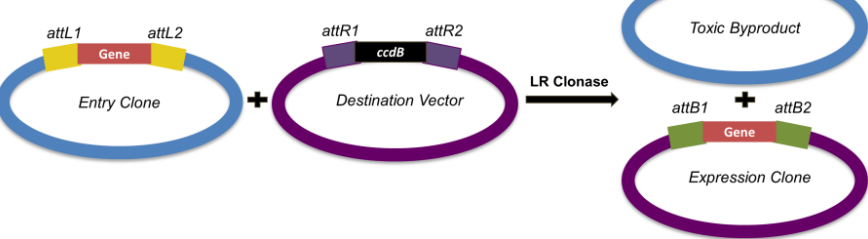
In Fusion-Creator cloning
fago P1 Cre-LoxP (Clontech)
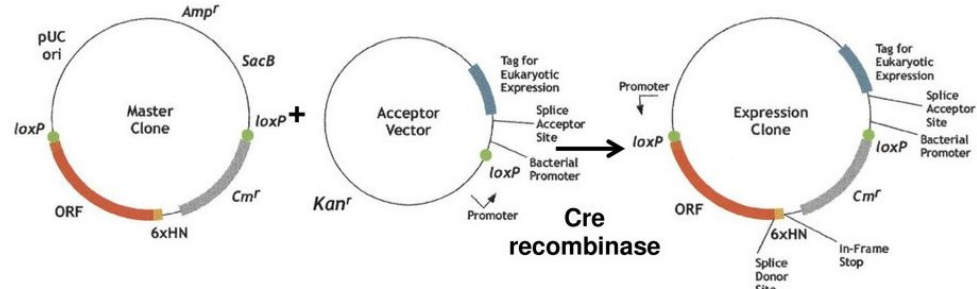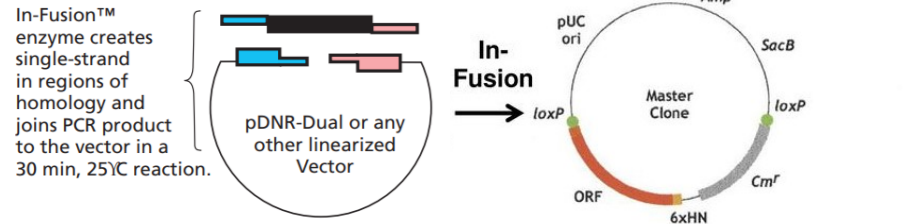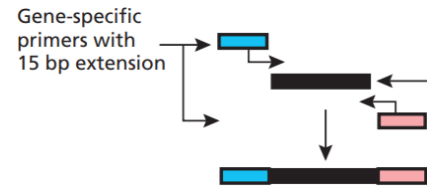
Espressione di proteine ricombinanti per studi strutturali
Inserimento nella cellula ospite:

Trasformazione

Batteri
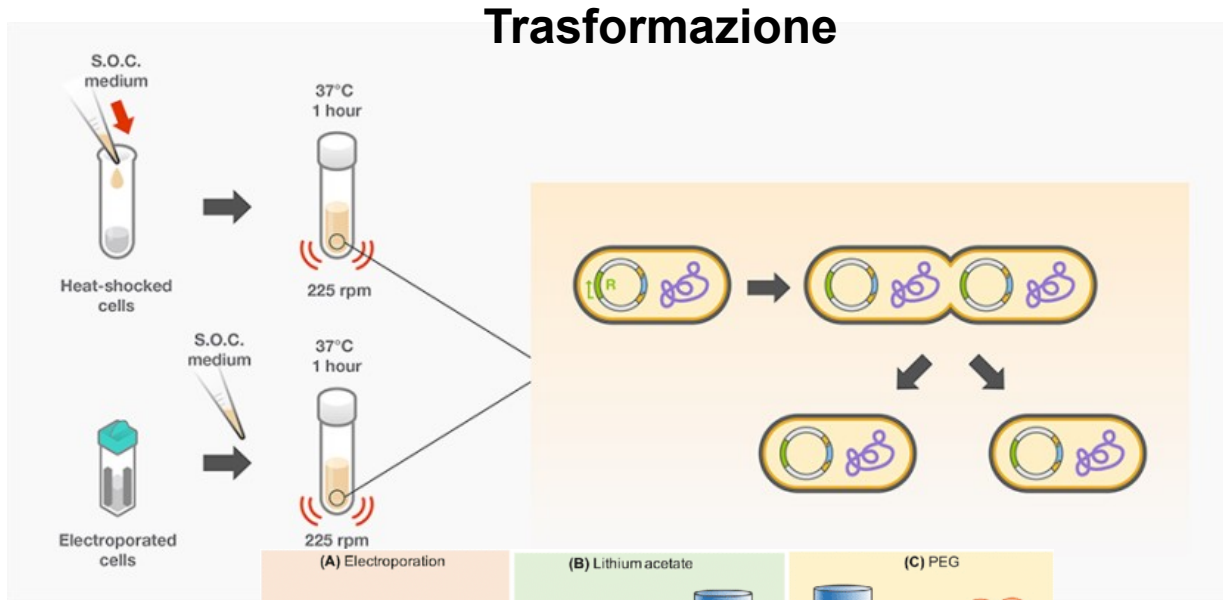
Lievito
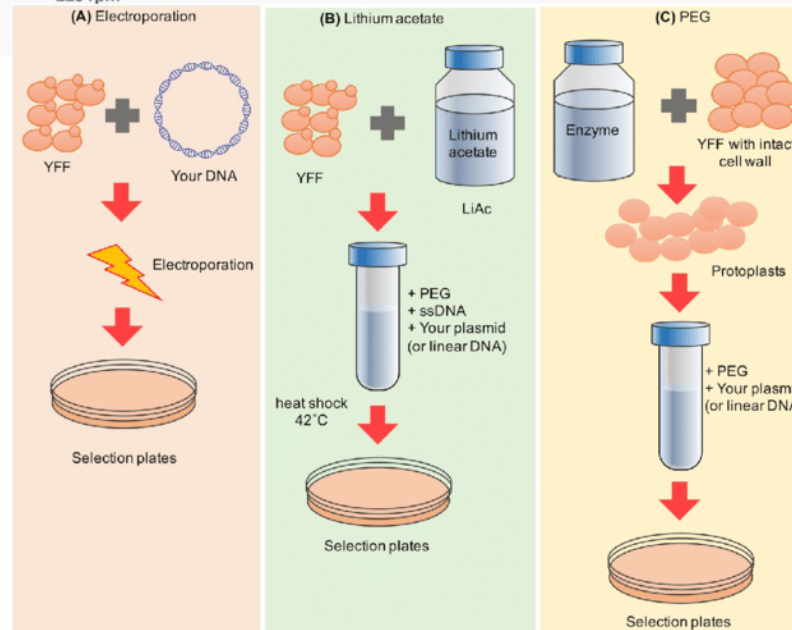
Espressione di proteine ricombinanti per studi strutturali
Inserimento nella cellula ospite:

Cellule insetto/mammifero:
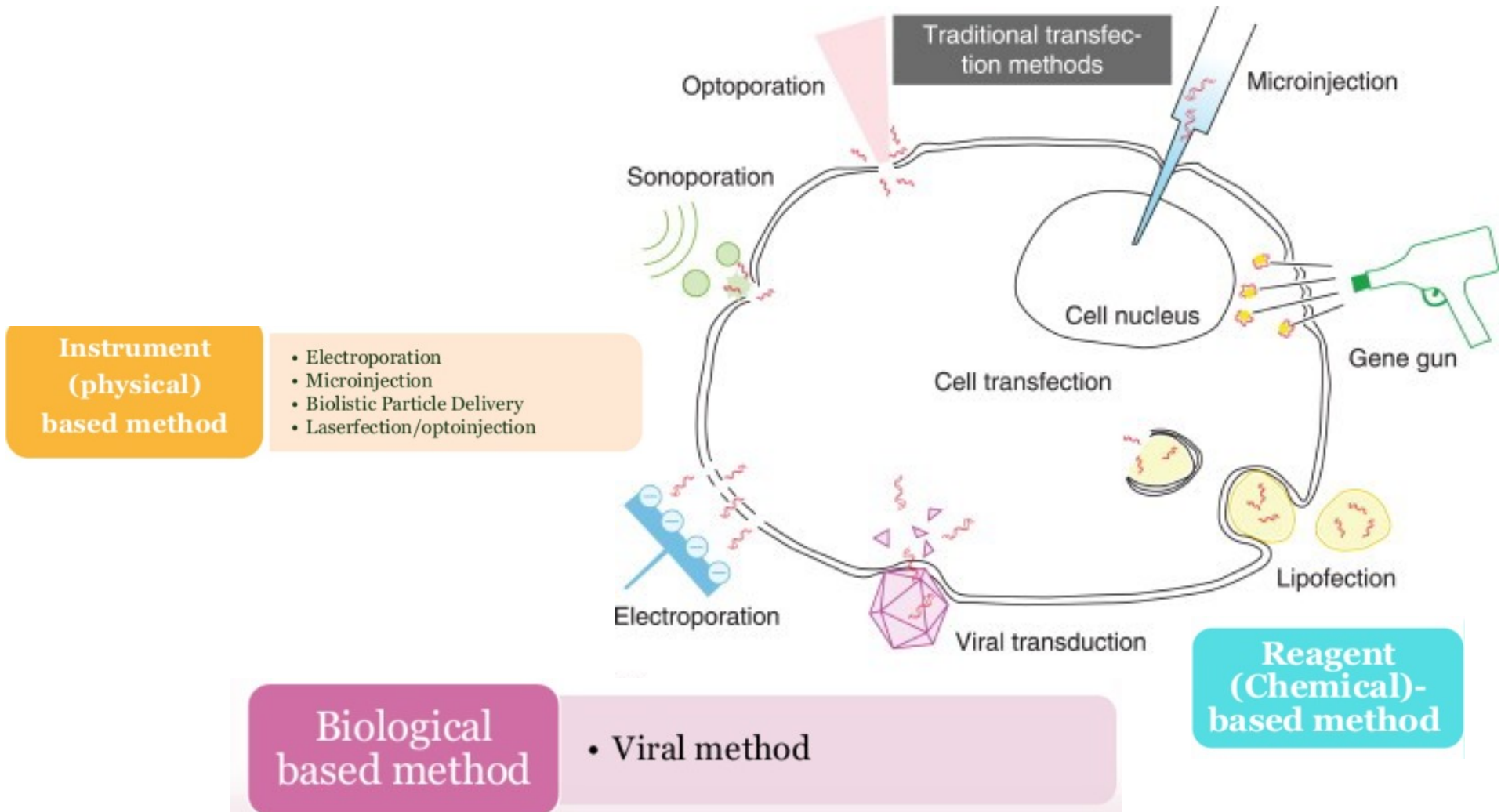
Espressione di proteine ricombinanti per studi strutturali
Inserimento nella cellula ospite:
Cellule insetto/mammifero:

Trasfezione
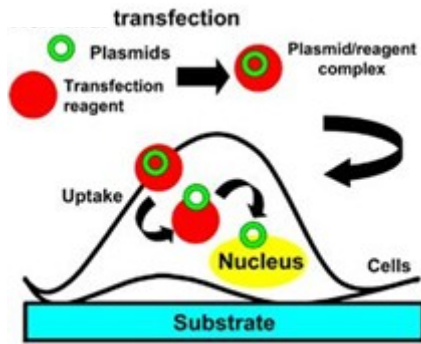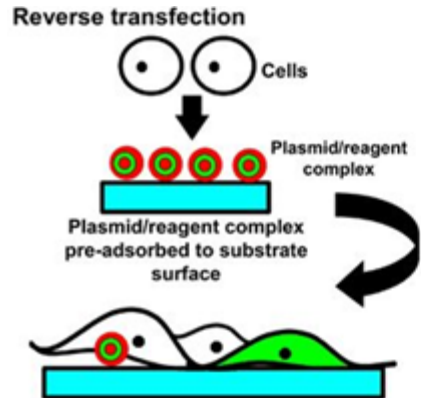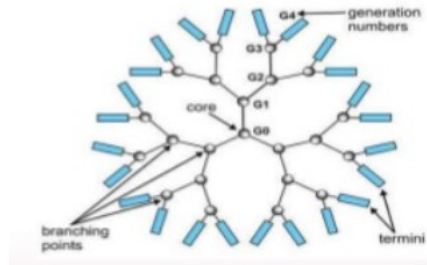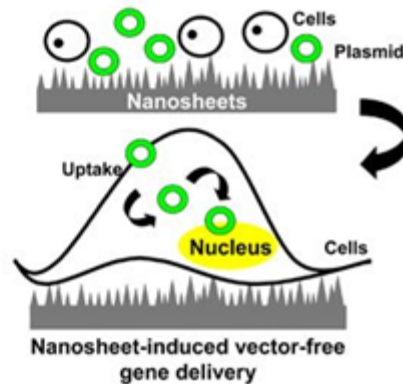
Trasduzione
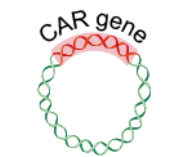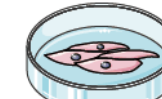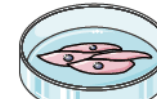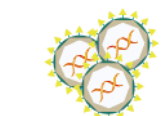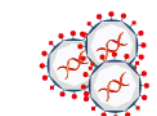
Espressione/purificazione di proteine ricombinanti per studi strutturali
Small scale inducible expression:

To test/optimize different strains/ expression conditions (°T/[inducer]/media/time)

EXPRESSION

HARVESTING

CELL LYSIS

CENTRIFUGATION

SDS-PAGE

TOTAL EXTRACTS

SOLUBLE FRACTIONS
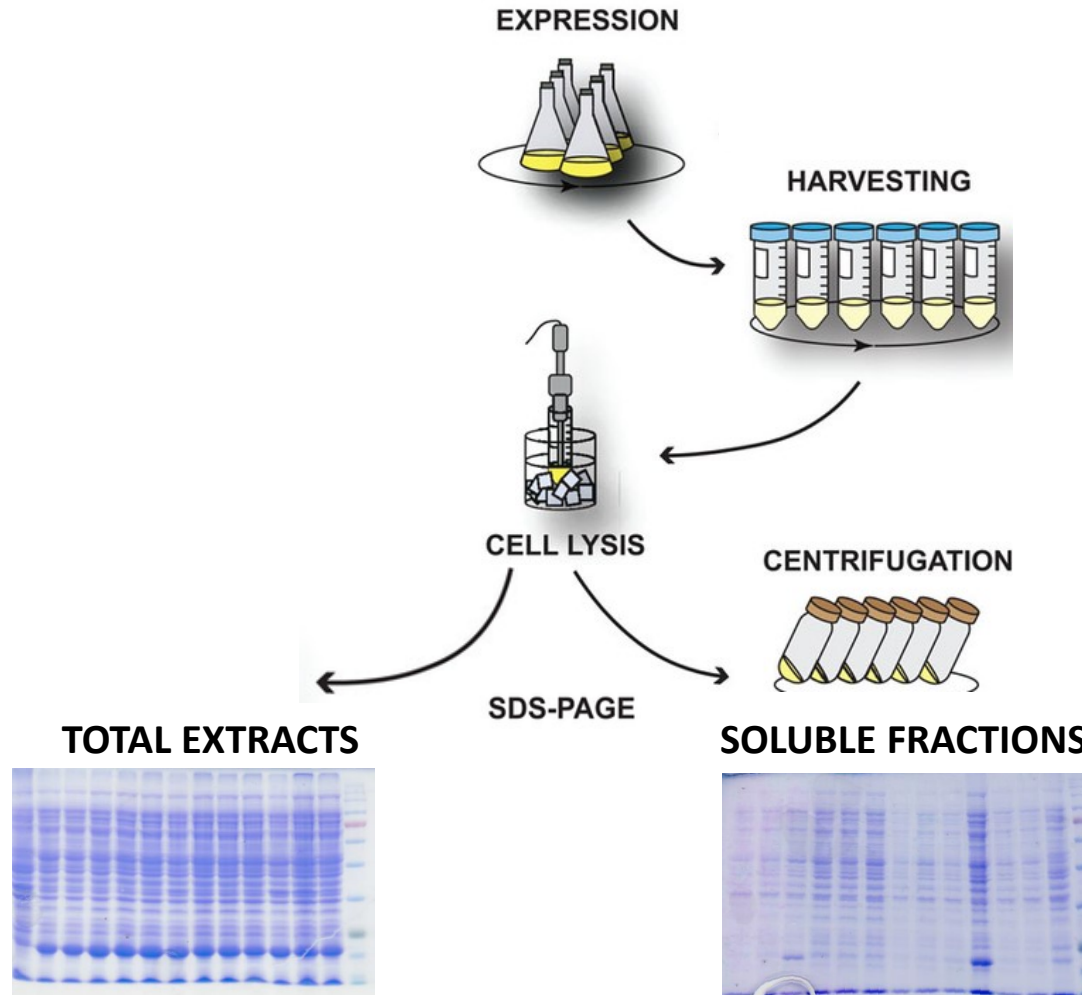
Large scale expression:
To adjust optimized expression conditions to larger scales

Espressione di proteine ricombinanti per studi strutturali
expression flowchart
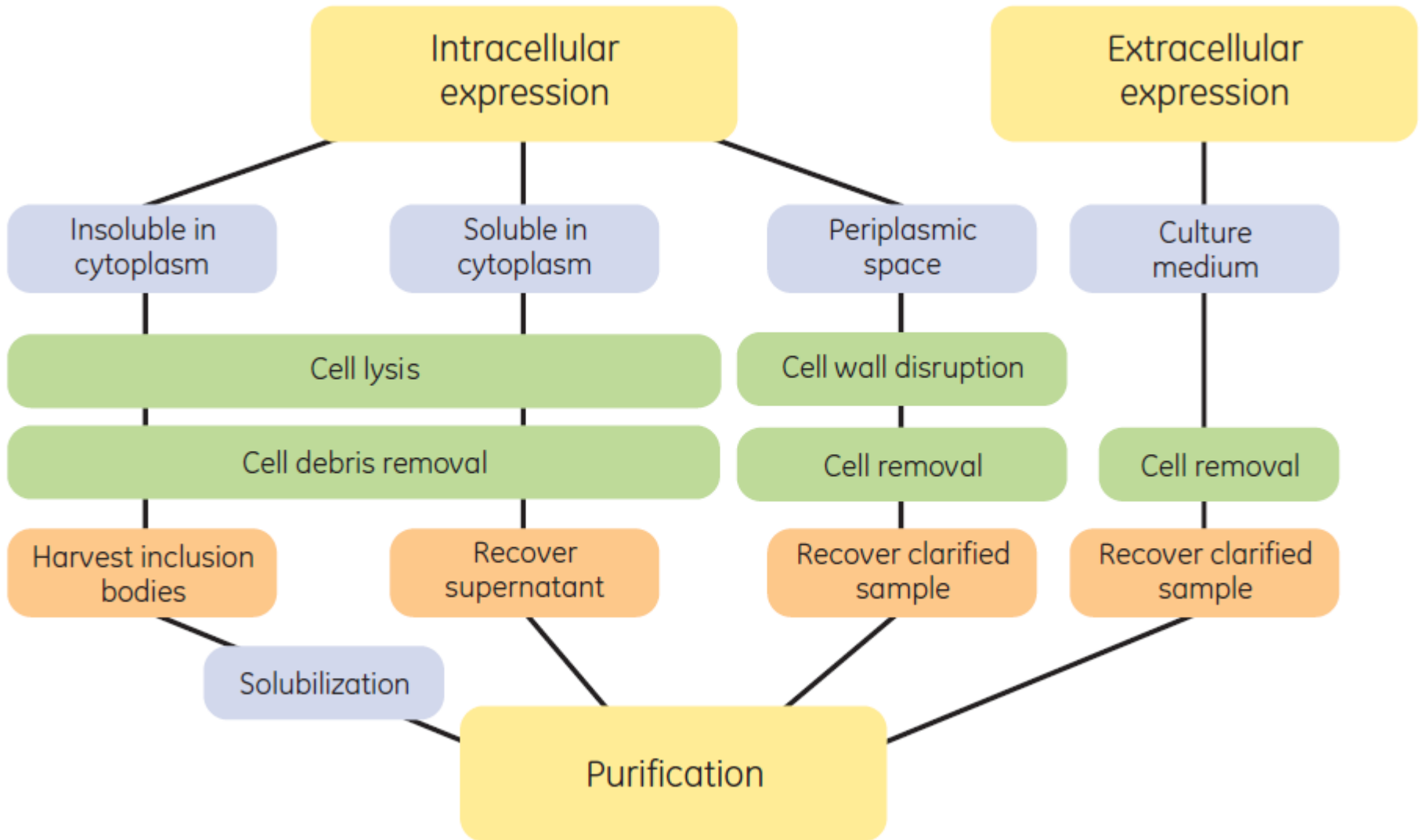
Espressione di proteine ricombinanti per studi strutturali
collection & lysis

CENTRIFUGAZIONE

      SURNATANTE per il recupero di proteine secrete

      PELLET: **LISI CELLULARE** per recuperare proteine citoplasmatiche / di membrana o proteine in corpi di inclusione lisato (proteina target, insieme a proteine solubili dell'ospite, membrane e proteine di membrana, organelli, DNA, matrice citosolica della cellula)

Cellule di insetto/MAMMIFERO: soluzione ipotonica (Shock osmotico)

Batteri e lieviti: più difficile

Metodo enzimatico (lisozima)

Sonicazione (alta pressione localizzata)

French press & Microfluidizer (valvola stretta ad alta pressione)
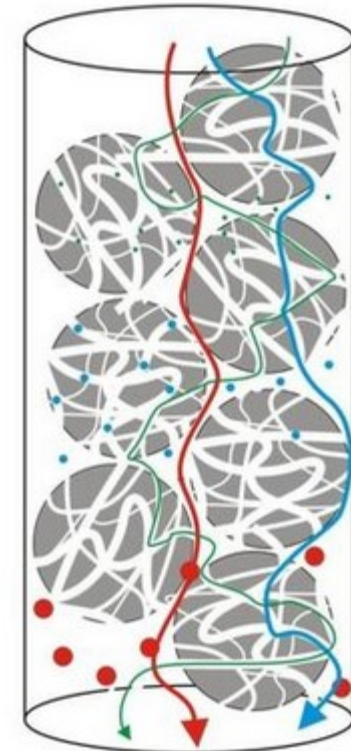
Bead Mill (microsfere e agitazione ad alta frequenza)
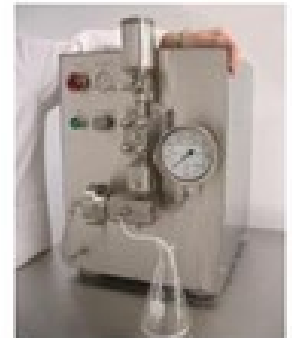
Purificazione di proteine ricombinanti per studi strutturali
purificazione

Purificazione di proteine ricombinanti per studi strutturali
purificazione
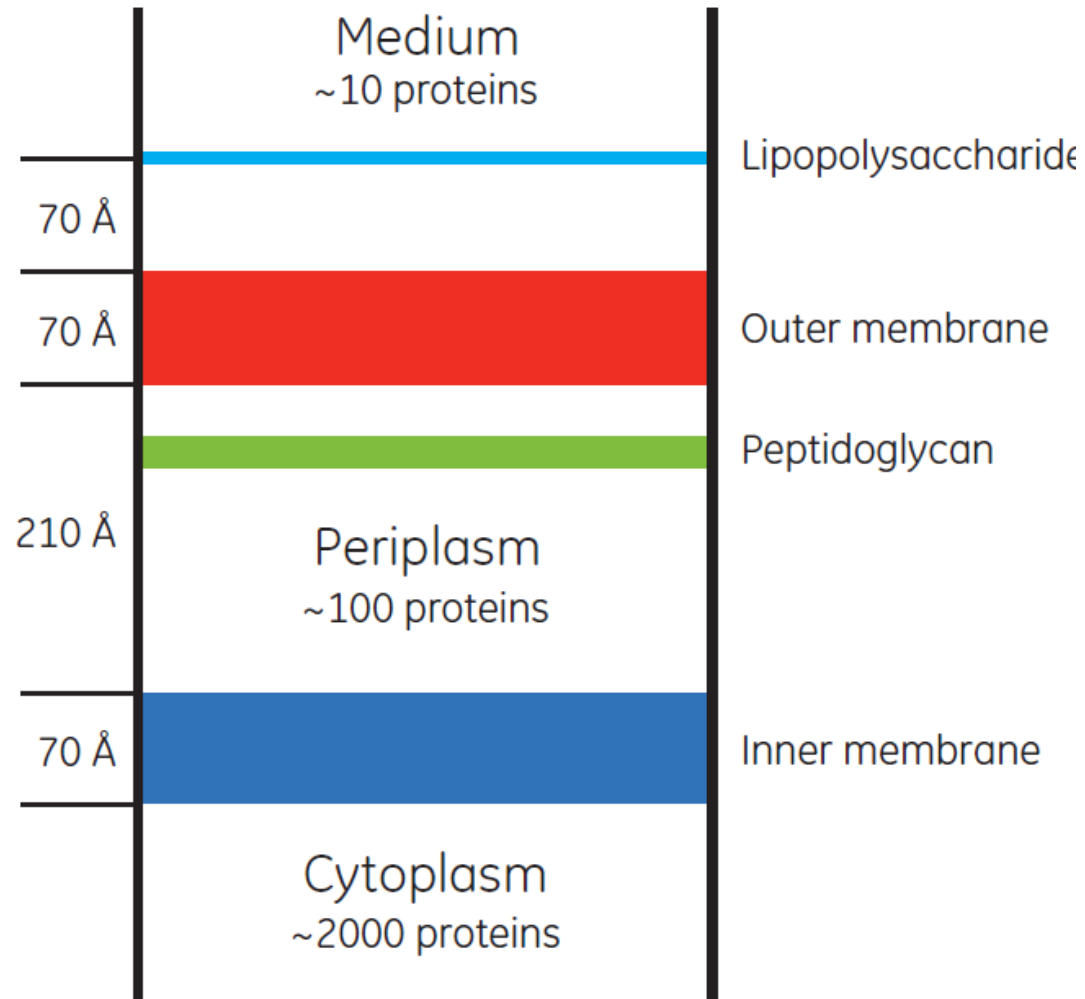
Purificazione di proteine ricombinanti per studi strutturali
evoluzione dei metodi di purificazione

Table 1. Some milestones in the history of protein purification

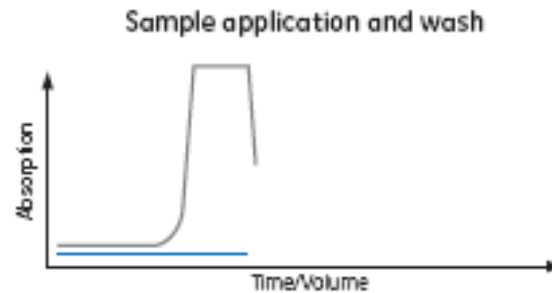|		
Precipitation	1789	Fourcroy
Crystalline ovalbumin	1889	Hofmeister
Chromatography	1903	Tswett*
Ultracentrifugation	1924	Svedberg
Moving boundary electrophoresis	1937	Tiselius
Ion exchange chromatography	1940s	The Manhattan Project
Two-phase partitioning	1955	Albertsson (1)
Size-exclusion chromatography (gel filtration)[†]	1955, 1959	Lindqvist and Storgårds (2), Porath and Flodin (3), Ingelman
Hydroxyapatite chromatography	1956	Tiselius, et al. (4)
Sephadex™ (gel filtration medium)	1959	Pharmacia (now GE Healthcare)
Polyacrylamide gel electrophoresis	1959	Raymond and Weintraub (5)
Isoelectric focusing[‡]	1959	Kolin (6)
High-performance liquid chromatography (HPLC)[§]	1941; 1966	Martin and Synge (7)
Sepharose™	1967	Pharmacia (now GE Healthcare)
SDS-PAGE	1967	Shapiro, et al. (8)
Affinity-ligand coupling chemistry for affinity chromatography[∥]	1967	Axén, et al. (9)
Affinity chromatography	1968	Cuatrecasas, et al. (10)
Reversed-phase chromatography	1970s	Kirkland (11); Molnar and Horvath (12)
Protein A adsorbent for IgG purification	1972	Hjelm, et al. (13); Kronvall (14)
Hydrophobic interaction chromatography	1973	Porath (15); Hjertén (16)
Two-dimensional chromatography	1975	O'Farrel l(17)
Immobilized metal ion affinity chromatography	1975	Porath, et al. (18)
Chromatofocusing	1977**	Sluyterman and Wijdenes (19)
Capillary electrophoresis	1981[††]	Jorgenson and Lukacs (20)
Fast Protein Liquid Chromatography	1982	Pharmacia (now GE Healthcare)
Histidine affinity tagging	1988[‡‡]	Smith et al. (21), Hochuli et al. (22)
ÄKTA™ design	1996	Pharmacia Biotech (now GE Healthcare)

Purificazione di proteine ricombinanti per studi strutturali
strategie di purificazione
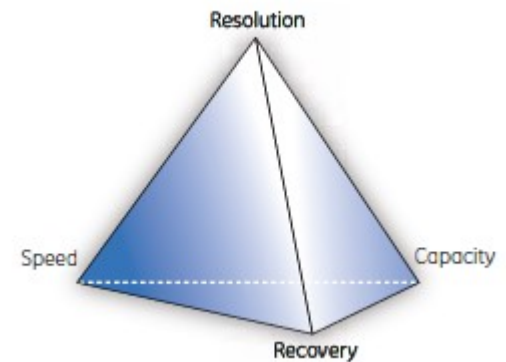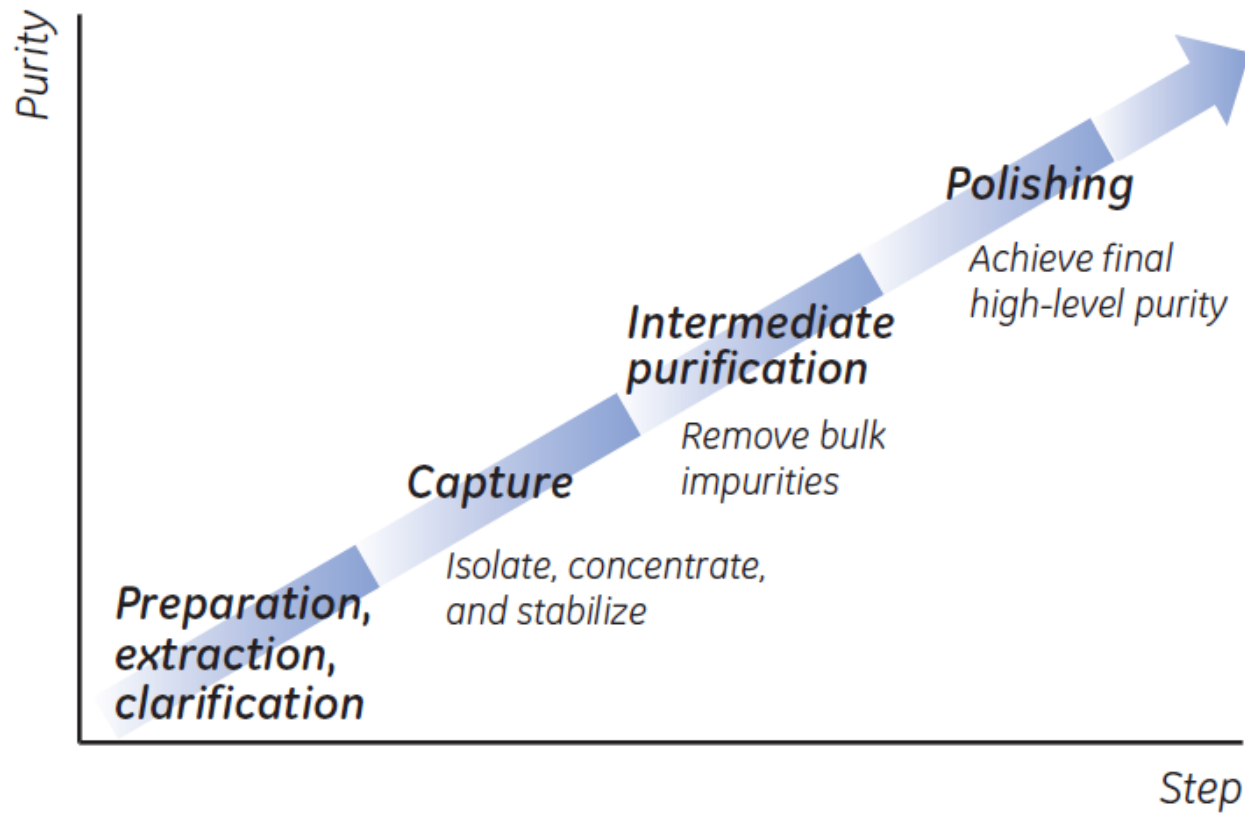
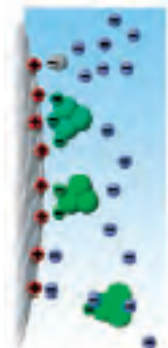# Purificazione di proteine ricombinanti per studi strutturali
strategie di purificazione: cattura
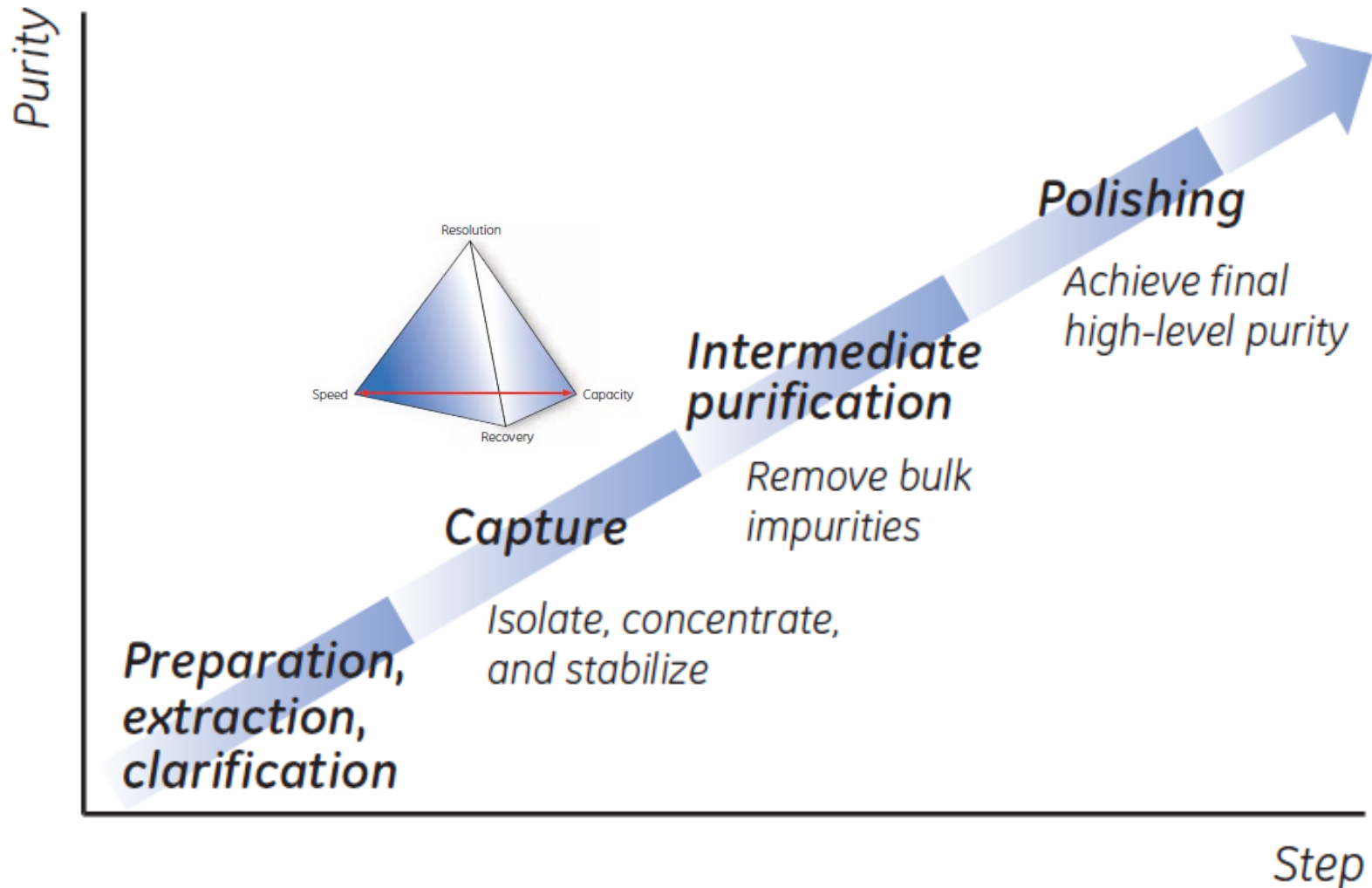
Purificazione di proteine ricombinanti per studi strutturali
strategie di purificazione: purificazione intermedia
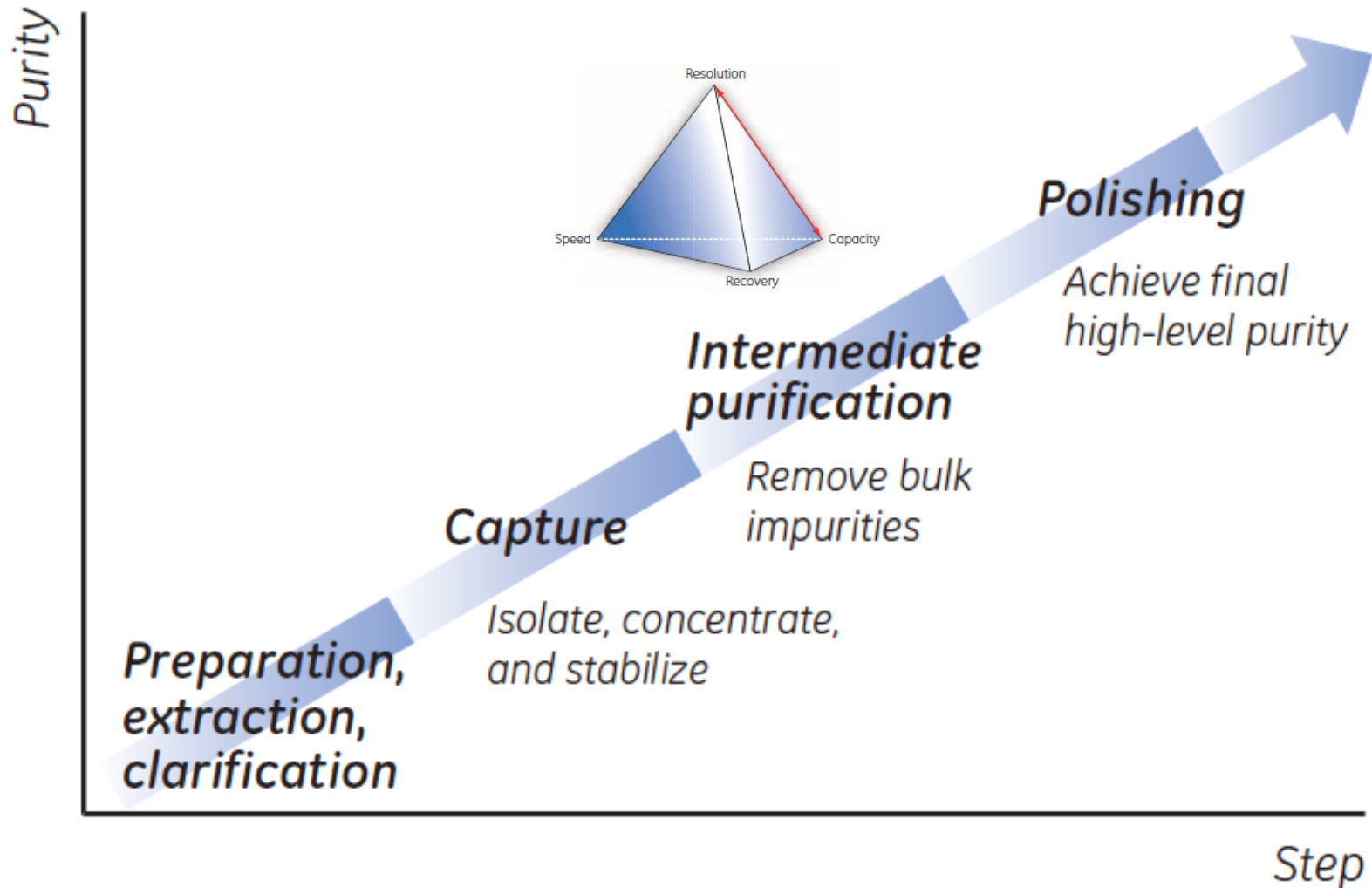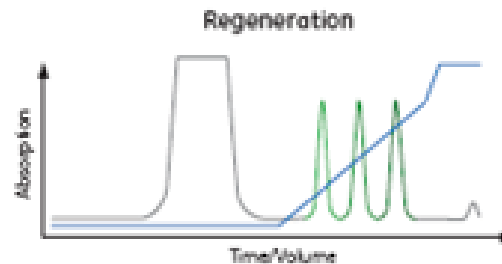
Purificazione di proteine ricombinanti per studi strutturali
strategie di purificazione: polishing
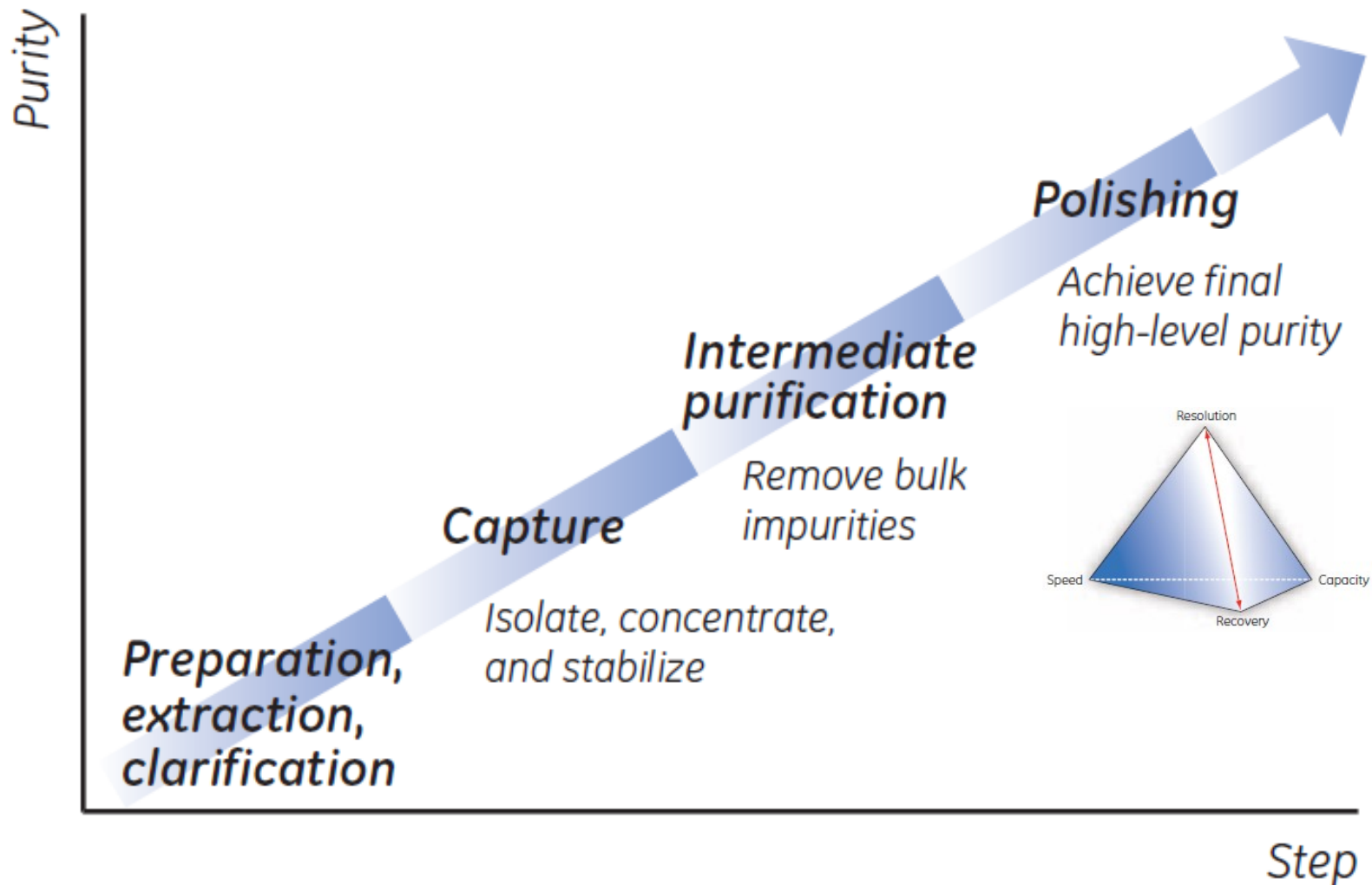
Purificazione di proteine ricombinanti per studi strutturali
Precipitazione frazionata:

separazione sulla base della solubilità (effetto del salting out)

Purificazione di proteine ricombinanti per studi strutturali
Precipitazione frazionata

Clarification
Bulk proteins and particulate matter precipitated

→ **Supernatant**

Extraction, Clarification, Concentration
Target protein precipitated with proteins of similar solubility

→ **Redissolve pellet***

Extraction, Clarification
Bulk proteins and particulate matter precipitated

→ **Concentration**
Target protein precipitated with proteins of similar solubility

→ **Redissolve pellet***
Remember: not all proteins are easy to redissolve, yield may be reduced

Purification
Remember: if precipitating agent is incompatible with next purification step, use Sephadex G-25 for desalting and buffer exchange, e.g., HiTrap Desalting, PD-10, or HiPrep 26/10 Desalting column (refer to Chapter 11)

Purificazione di proteine ricombinanti per studi strutturali
Precipitazione frazionata

Precipitation agent	Typical conditions	Sample type	Comment
Ammonium sulfate	As described later in this chapter.	> 1 mg/ml proteins, especially immunoglobulins.	Stabilizes proteins, no denaturation. Supernatant can go directly to HIC. Helps to reduce lipid content.
Dextran sulfate	Add 0.04 ml of 10% dextran sulfate and 1 ml of 1 M CaCl$_2$ per ml of sample, mix 15 min, centrifuge at 10 000 × g, discard pellet.	Samples with high levels of lipoproteins, e.g., ascites.	Precipitates lipoproteins.
Polyvinylpyrrolidone	Add 3% (w/v), stir 4 h, centrifuge at 17 000 × g, discard pellet.	Samples with high levels of lipoprotein, e.g., ascites.	Alternative to dextran sulfate.
Polyethylene glycol (PEG, M$_r$ > 4000)	Up to 20% (w/v).	Plasma proteins.	No denaturation, supernatant goes directly to IEX or AC. Complete removal of PEG may be difficult. Stabilizes proteins.
Acetone (cold)	Up to 80% (v/v) at 0°C. Collect pellet after centrifugation at full speed in chilled microcentrifuge.	Useful for peptide precipitation or concentration of sample for electrophoresis.	May denature protein irreversibly.
Polyethyleneimine	0.1% (w/v).		Precipitates aggregated nucleoproteins.
Protamine sulfate	1% (w/v).		Precipitates aggregated nucleoproteins.
Streptomycin sulfate	1% (w/v).		Precipitates nucleic acids.
Caprylic acid	1:15 (w/w).	Antibody concentration should be > 1 mg/ml.	Precipitates bulk of proteins from sera or ascites, leaving immunoglobulins in solution.

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:
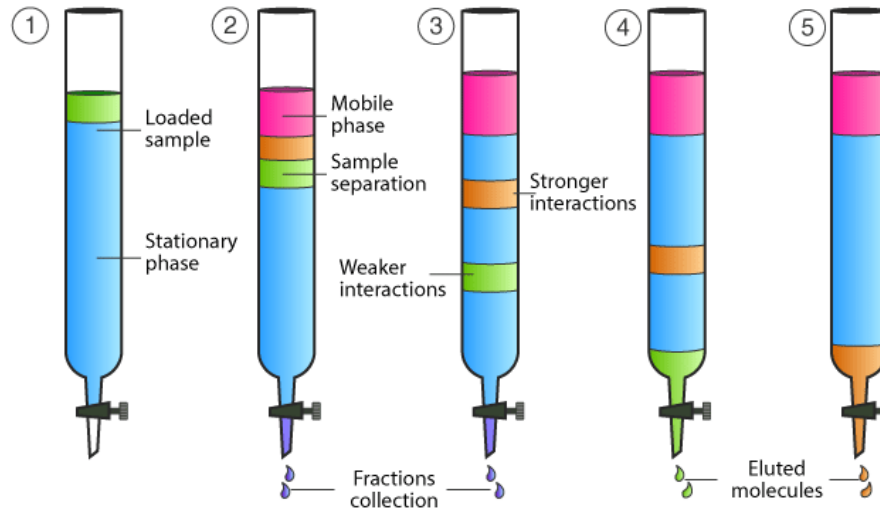
IN BATCH

CON FPLC (Fast Protein Liquid Chromatography)
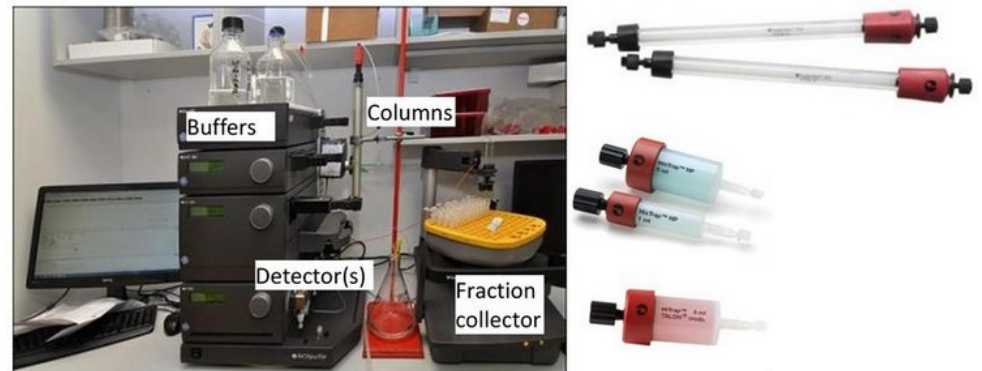
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:
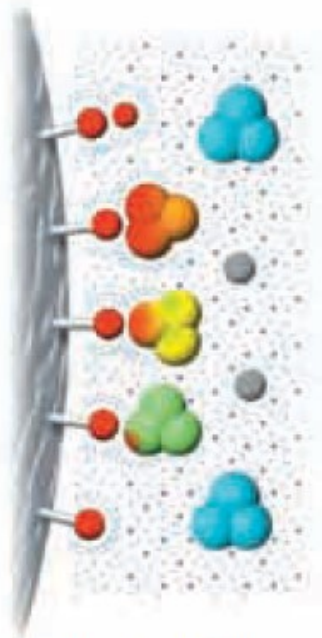
Gel filtration   Hydrophobic interaction   Ion exchange   Affinity   Reversed phase
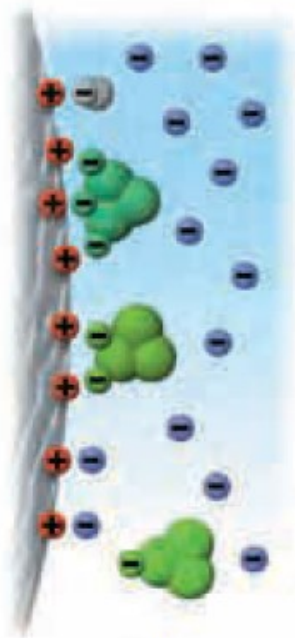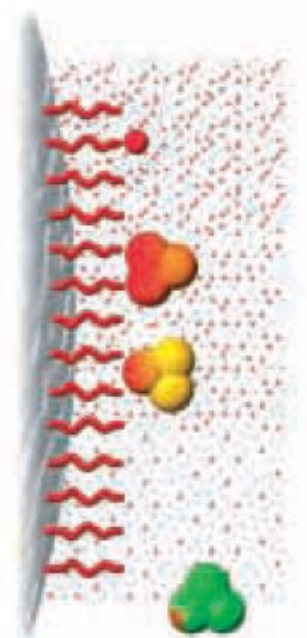
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici

SEC
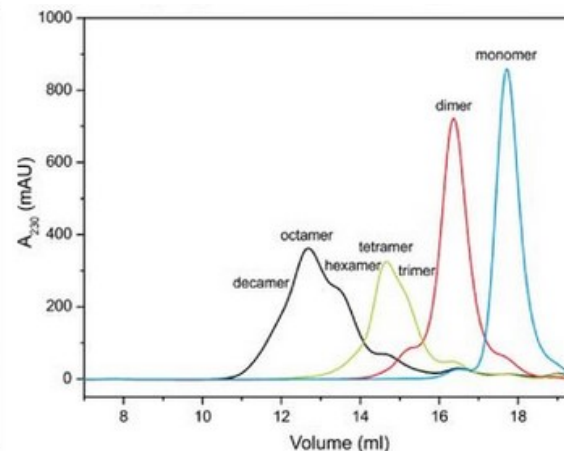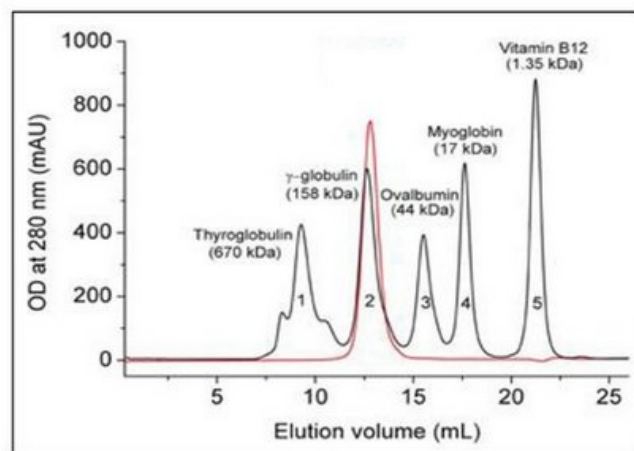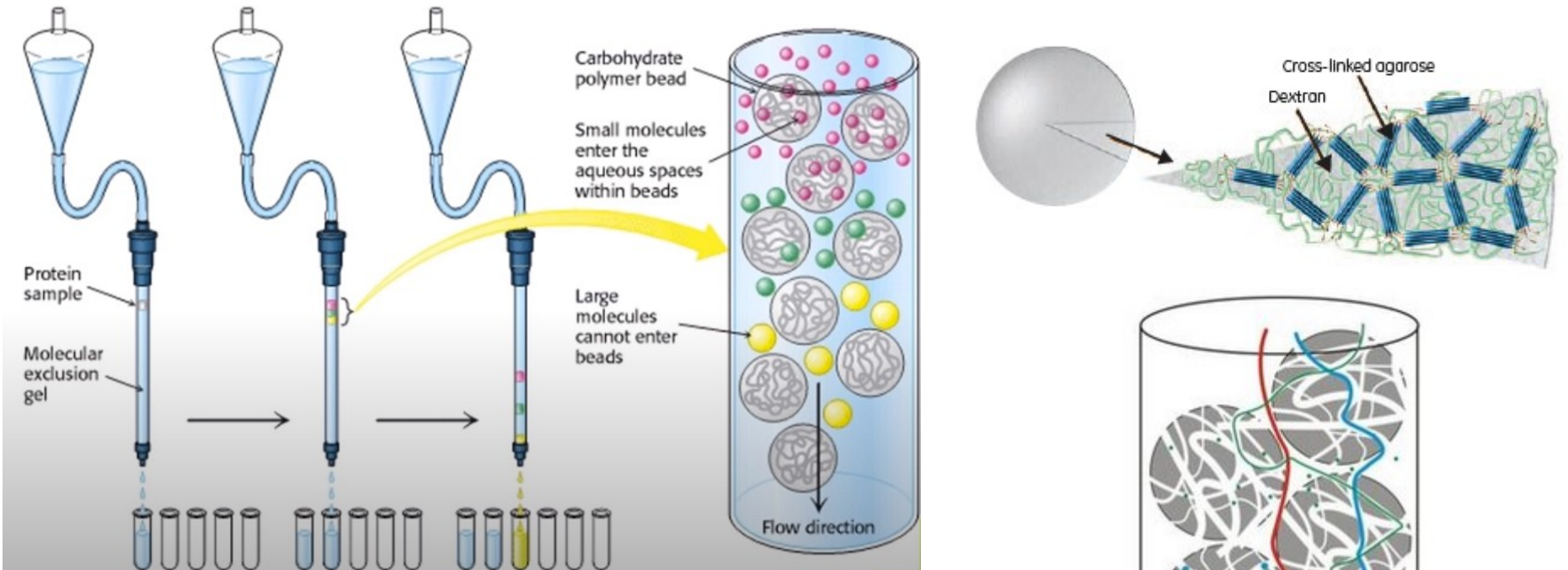
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici

SEC
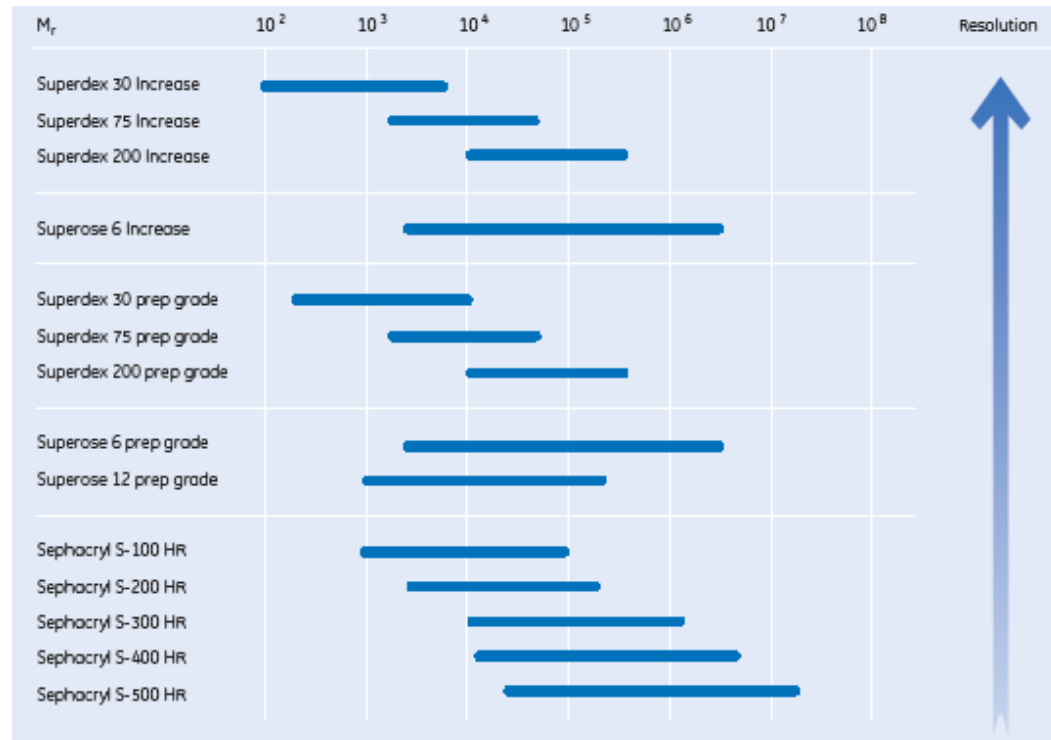
Product	Efficiency[1]	Operational[2]	Cleaning-in-place (CIP)[3]	Particle size, d_{50v} (µm)[4]
Superdex 30 Increase	> 43 000	3 to 12	1 to 14	~ 9
Superdex 75 Increase	> 43 000	3 to 12	1 to 14	~ 9
Superdex 200 Increase	> 48 000	3 to 12	1 to 14	~ 8.6
Superdex 30 prep grade	≥ 13 000	3 to 12	1 to 14	~ 34
Superdex 75 prep grade	≥ 13 000	3 to 12	1 to 14	~ 34
Superdex 200 prep grade	≥ 13 000	3 to 12	1 to 14	~ 34

Purificazione di proteine ricombinanti per studi strutturali
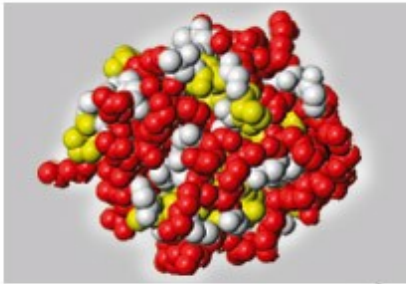## Metodi cromatografici

IEX
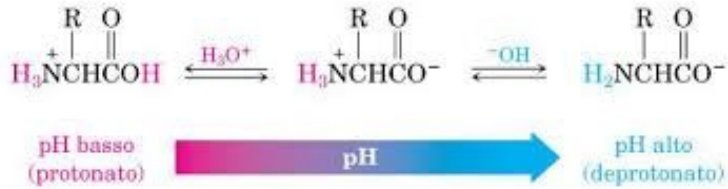
CATIONICO   Punto isoelettrico (zwitterione neutro) (pI 5.5-7.5)   **ANIONICO**

Resin	Functional group	Weak or Strong	Functional pH Range
DEAE	diethylaminoethyl [-N⁺(C₂H₅)₂H⁺]	weak anion	pH 2 - 9
ANX	diethylaminopropyl [-N⁺(C₂H₅)₂H⁺]	weak anion	pH 2 - 9
Q	quarternary amine [-N⁺(CH₃)₃]	strong anion	pH 1 - 14
CM	carboxymethyl [-O-CH₂COO⁻]	weak cation	pH 5 - 10
S	methyl sulfonate [O-CH₂-CHOH-CH₂-O-CH₂-CHOH-CH₂-SO₃⁻]	strong cation	pH 2 - 12
SP	sulphonyl group [-CH₂-CH₂-CH₂SO₃⁻]	strong cation	pH 2 - 14

Tertiary amines ⟶ Strong anion exchangers
Secondary amines ⟶ Weak anion exchangers

Sulphonic acid ⟶ Strong cation exchangers
Carboxylic acid ⟶ Weak cation exchangers

- Positively charged exchangers
- Mobile counter ion
- Ion exchanger resin

- Negatively charged exchangers
- Mobile counter ion
- Ion exchanger resin

- selettività diversa
- capacità varia con il pH per i deboli

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici

IEX: fasi della purificazione

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici

IEX: fasi della purificazione

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici

IEX: dimensioni delle componenti di matrice

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici
IEX: selettività vs efficienza

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici
IEX: tipi di eluizione
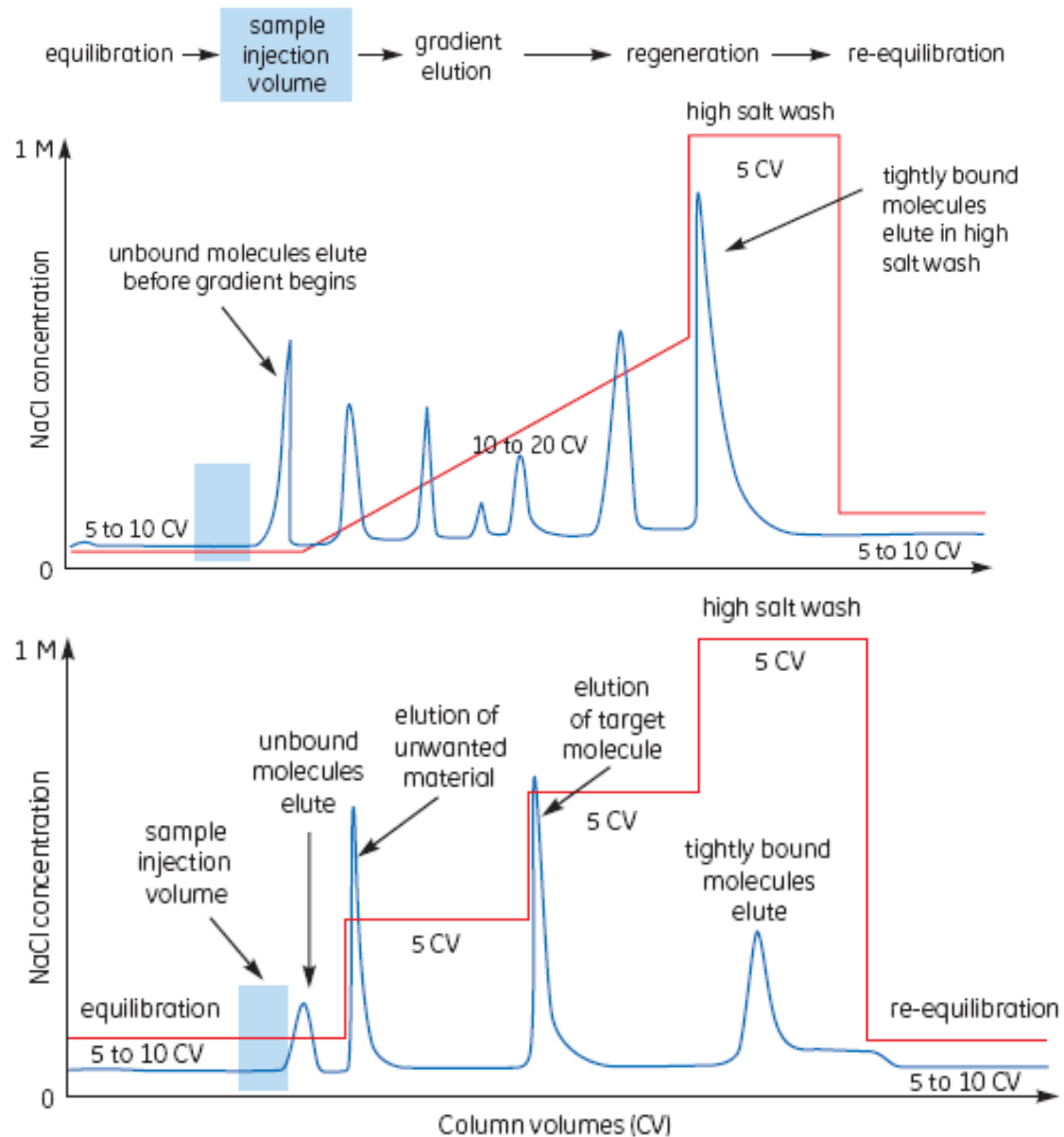
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

IEX: CAPTURE

Capture
Isolate, concentrate, and stabilize target protein(s)
Sample conditions: clarified or nonclarified

Easy scale-up, Broad choice of selectivity, including alternatives to Q or S IEX media
Sepharose Fast Flow
(Q, SP, DEAE, CM, ANX)

Try weak ion exchangers such as DEAE, CM, or ANX if the selectivity of Q or S is unsatisfactory.

High volume throughput and high capacity, Easy scale-up
Capto (Q, S, DEAE, adhere, MMC)

Use high bed heights for increased productivity. Use Capto MMC for high salt feed.

High binding capacity for selected proteins, Easy scale-up
Sepharose XL (Q or SP)

Use for capture with high binding capacity/rapid separation of proteins from clarified samples

Large scale, viscous samples
Sepharose Big Beads (Q or SP)

Use with step elution.

Q Sepharose Fast Flow
(HR 16 × 50 mm)

Sample:
Pancreatin
Gradient elution

Q Sepharose XL
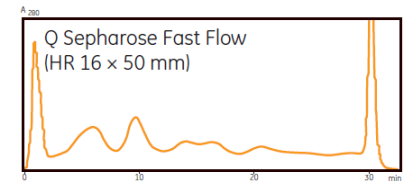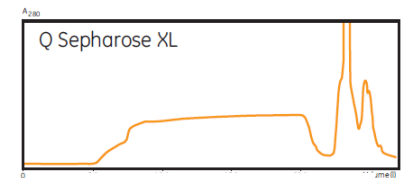
Sample:
Recombinant α-amylase
Pilot scale:
Gradient elution begins after 20 l

RISOLUZIONE

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

IEX: INTERMEDIATE PURIFICATION

Intermediate purification
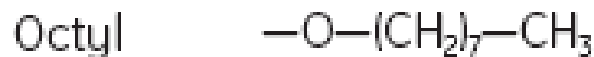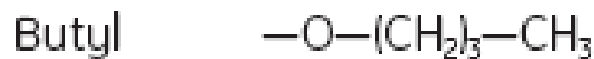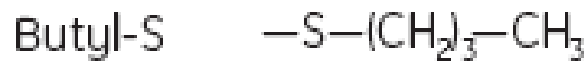Remove bulk impurities
Sample condition:
partially purified

High resolution
Easy scale-up
Sepharose High Performance (Q or SP)

Use HiTrap columns prepacked with Sepharose High Performance, Sepharose XL and Sepharose Fast Flow for media selection and pH scouting.

High resolution and throughput
Flexibility of process design
Capto ImpRes (Q or SP)
Capto S ImpAct

Use Capto ImpRes and Capto S ImpAct for increased productivity in large-scale production

Q Sepharose High Performance (HR 16 × 50 mm)
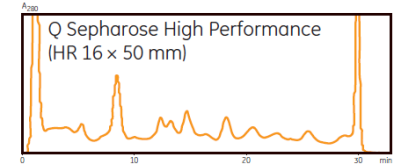
Sample:
Pancreatin
Gradient elution

RISOLUZIONE

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

IEX: POLISHING
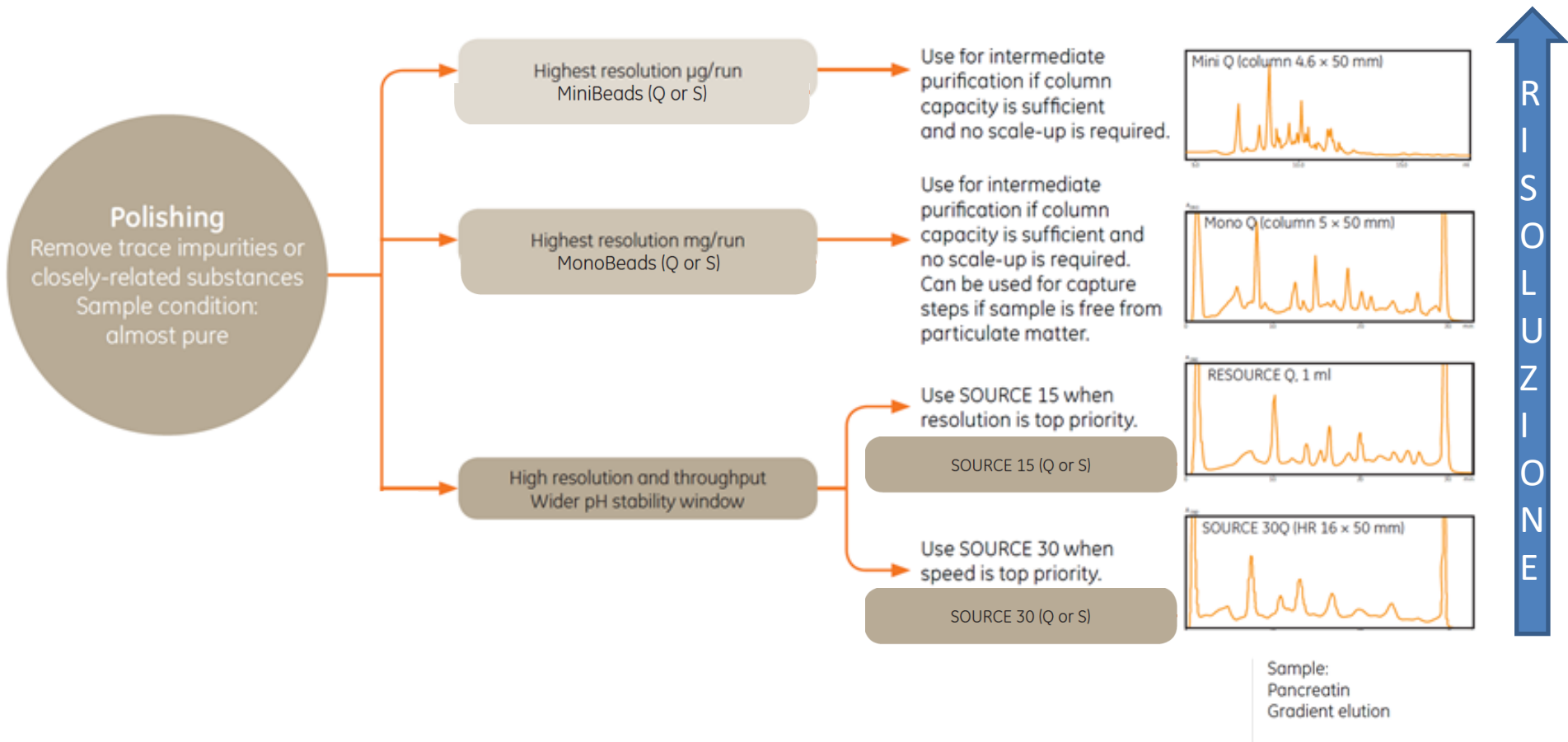
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

HIC
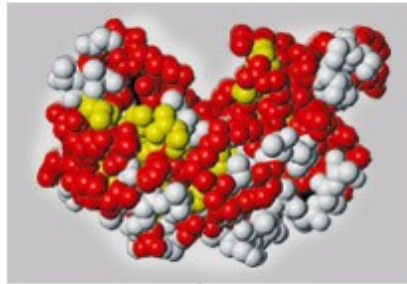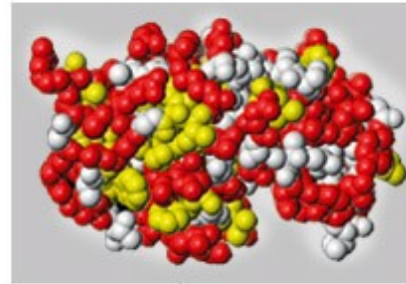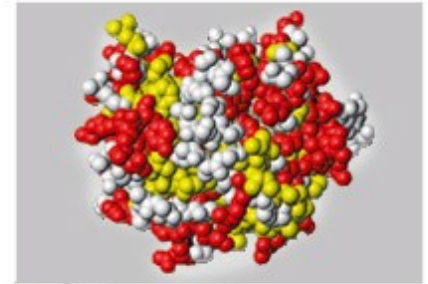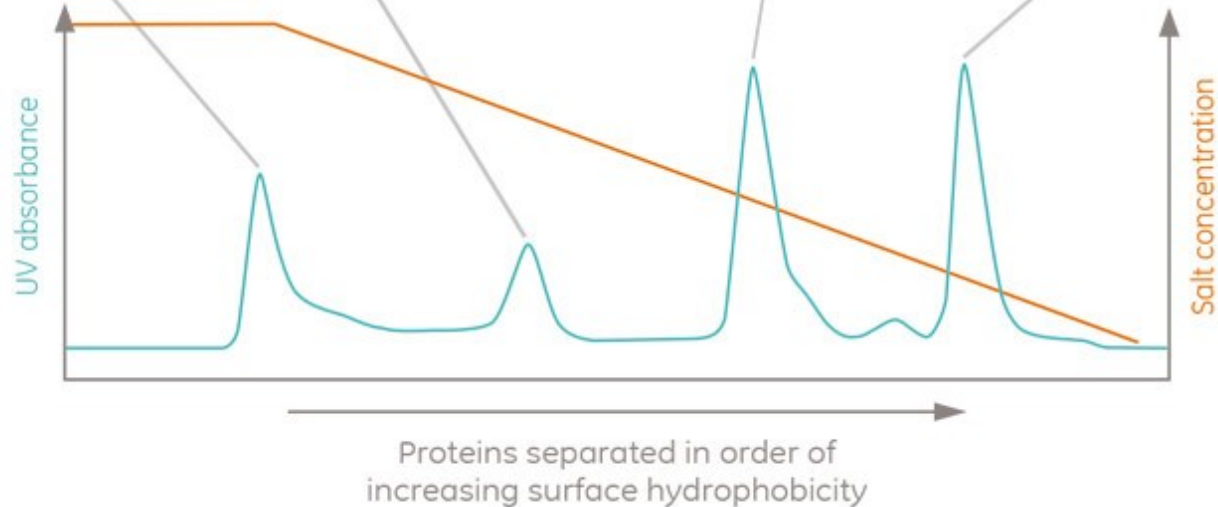
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

HIC
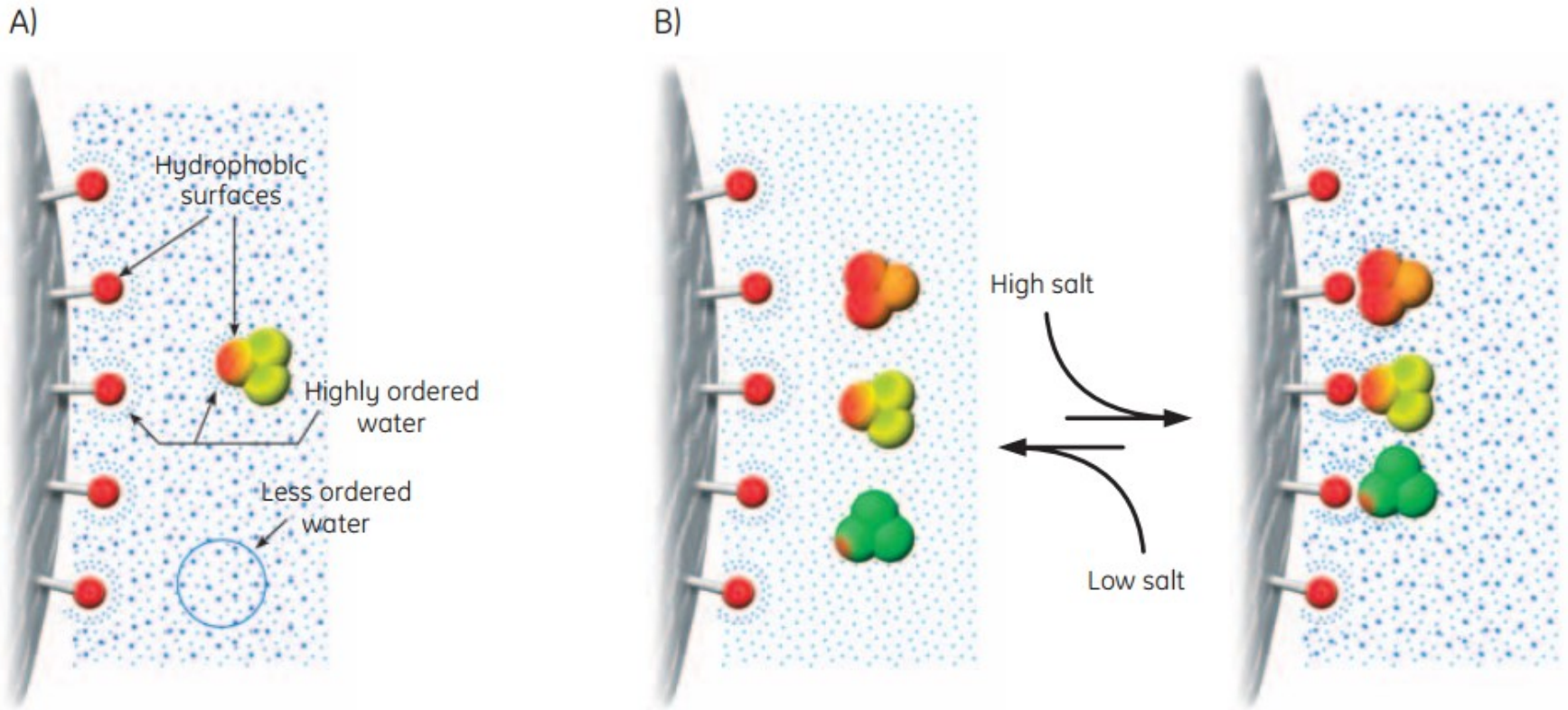
A) Highly ordered water shells surround the hydrophobic surfaces of ligands and proteins. Hydrophobic substances are forced to merge to minimize the total area of such shells (maximize entropy). Salts enhance the hydrophobic interaction. B) The equilibrium of the hydrophobic interaction is controlled predominantly by the salt concentration.

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

HIC: fasi della purificazione

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

HIC: fasi della purificazione

Elution 2
Further decreases in salt displace the more hydrophobic proteins (more tightly bound).
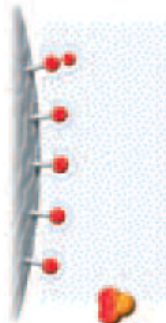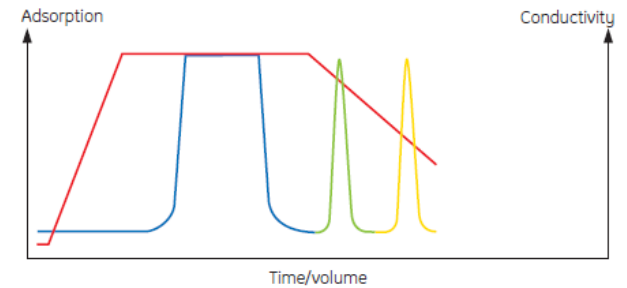
Elution 3
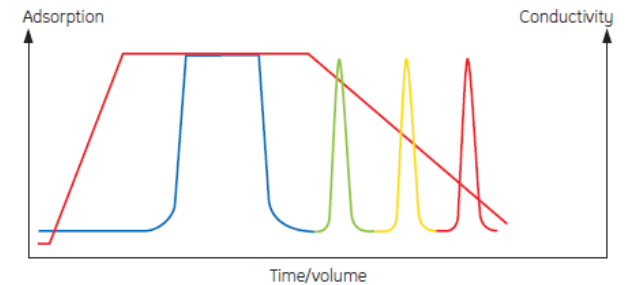
Wash
Final "salt-free" wash removes any hydrophobically bound proteins before re-equilibration.
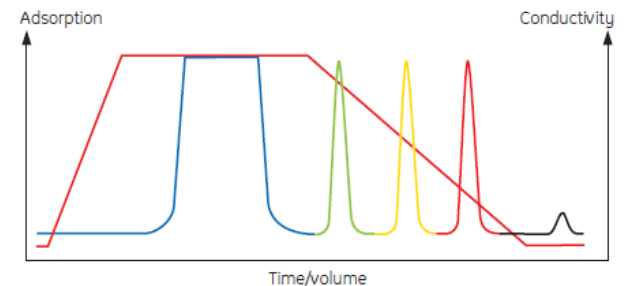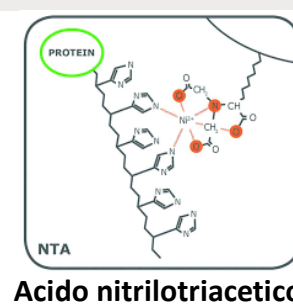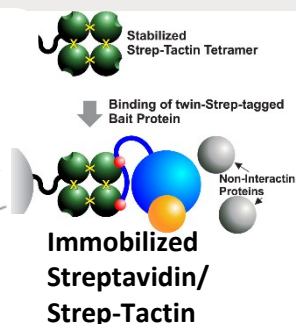
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

HIC: tipi di matrici

Matrix	Form	Mean particle size
SOURCE 15	Polystyrene/divinyl benzene	15 µm
Sepharose High Performance	Agarose 6%	34 µm
Sepharose 6 Fast Flow	Agarose 6%	90 µm
Sepharose 4 Fast Flow	Agarose 4%	90 µm
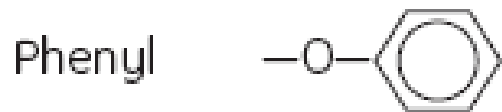
Phenyl —O—⬡

Butyl-S —S—$(CH_2)_3$—CH_3

Butyl —O—$(CH_2)_3$—CH_3

Octyl —O—$(CH_2)_7$—CH_3

Ether —O—CH_2—CHOH—CH_2—OH

Isopropyl —O—CH—$(CH_3)_2$

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

HIC: tipi di eluizione
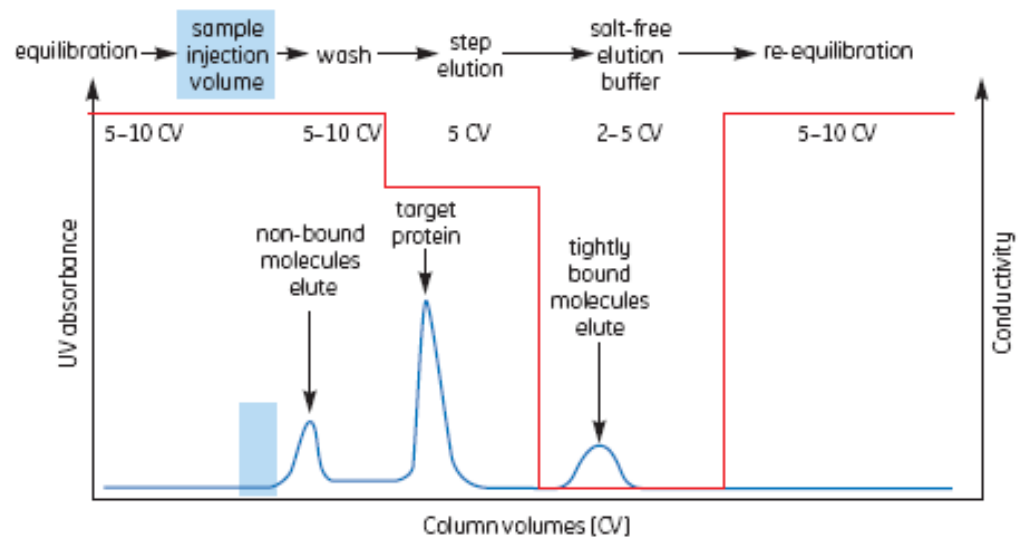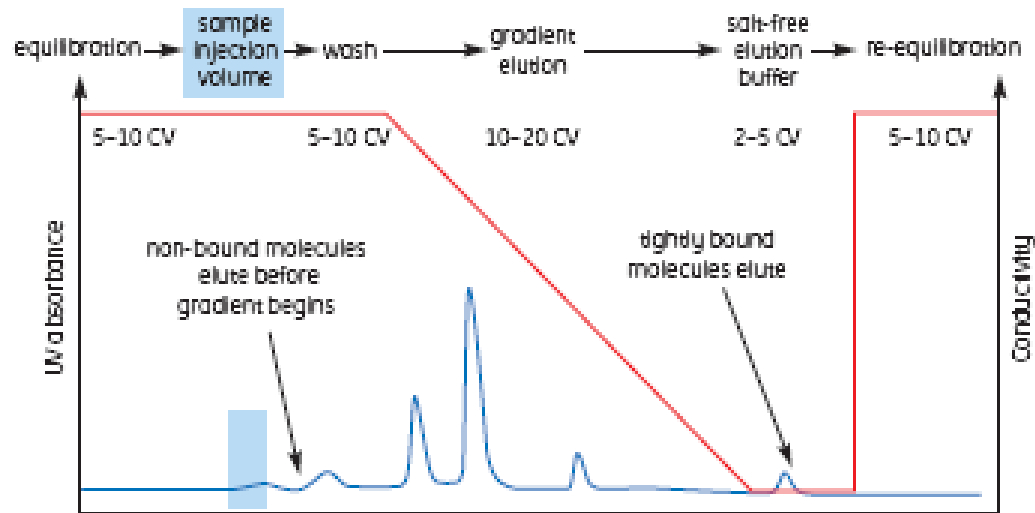
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

HIC: CAPTURE INTERMEDIATE PURIFICATION

Intermediate purification

Remove bulk impurities
Sample condition:
partially purified

Capture

Isolate, concentrate
and stabilize target
protein(s)
Sample condition:
clarified or
non-clarified

Good resolution
Easy scale-up
Range of selectivities

Sepharose Fast Flow
(phenyl high sub,
phenyl low sub, butyl,
octyl, butyl-S)

Use Sepharose Fast Flow
for polishing if media with
smaller particle size do not
offer required selectivity.

Good resolution
Industrial-scale,
filtration and capture
in one step

STREAMLINE
(Phenyl)
available as a CDM product

Use STREAMLINE for direct
capture from unclarified
feedstock.

Phenyl Sepharose 6 Fast Flow (high sub)

RISOLUZIONE

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

HIC: INTERMEDIATE PURIFICATION & POLISHING
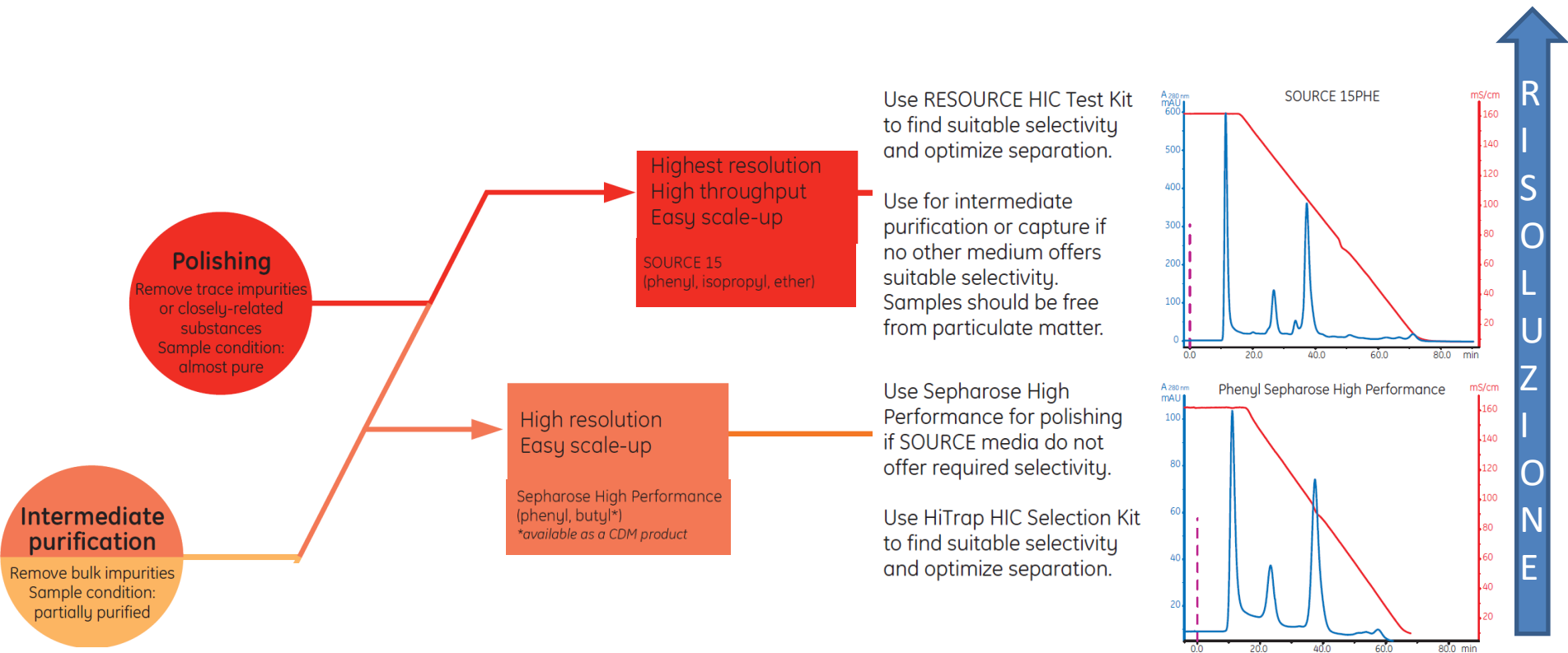
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:
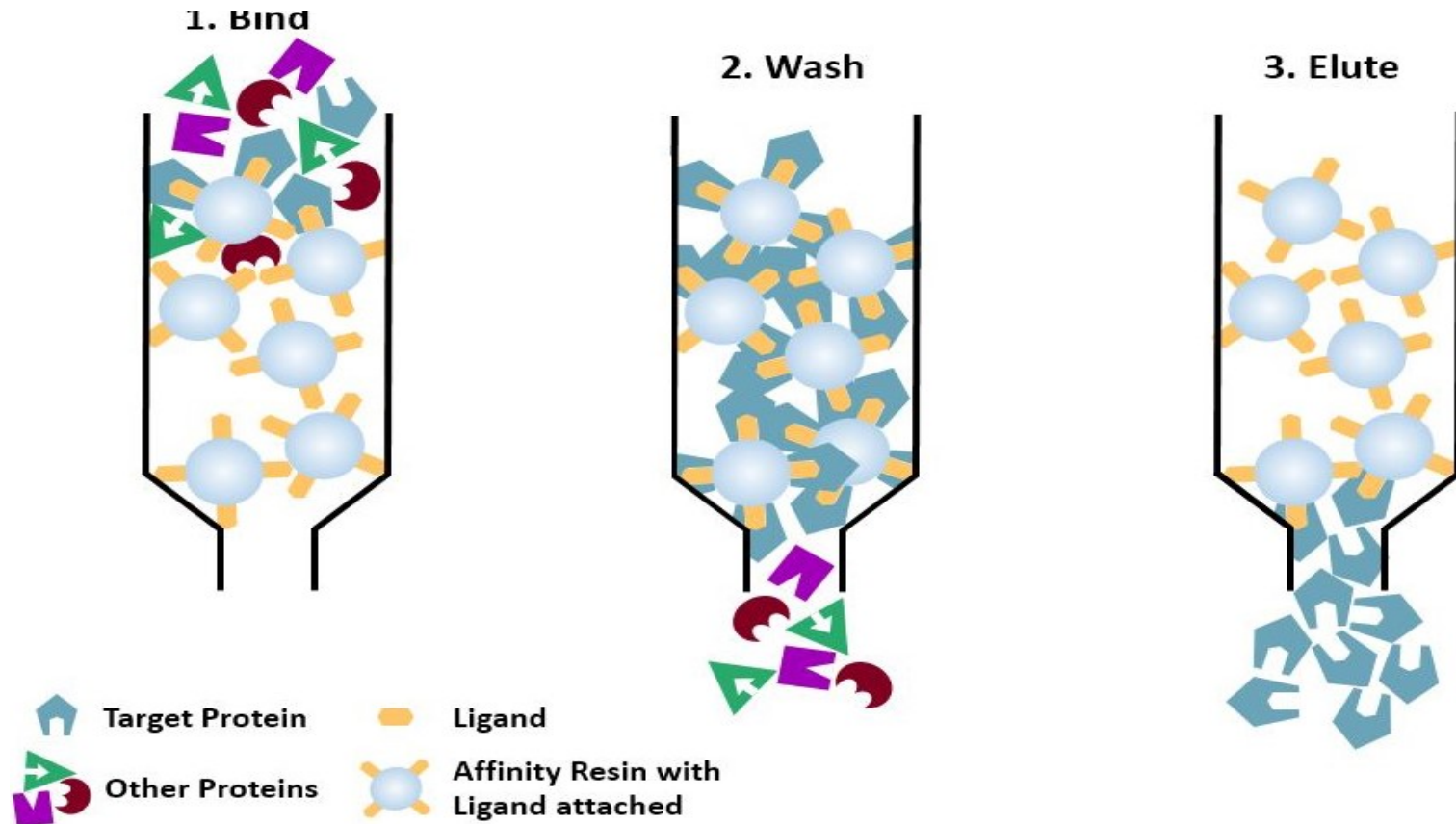
AFFINITY

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

**AFFINITY**

Types of ligand	Target molecules or molecules of interest
Enzyme	Substrate analogue, inhibitor, cofactor
Antibody	Antigen
Lectin	Polysaccharide, glycoprotein, cell surface receptor, cell
Nucleic acid	Complementary base sequence, nucleic acid binding protein
Hormone	Receptor
Avidin	Biotin
Calmodulin	Calmodulin-binding molecule
Poly(A)	RNA containing poly(U) sequences
Glutathione	Glutathione-S-transferase or GST fusion proteins
Proteins A and G	Immunoglobulins
Metal ions	Poly (His) fusion proteins, native proteins with histidine
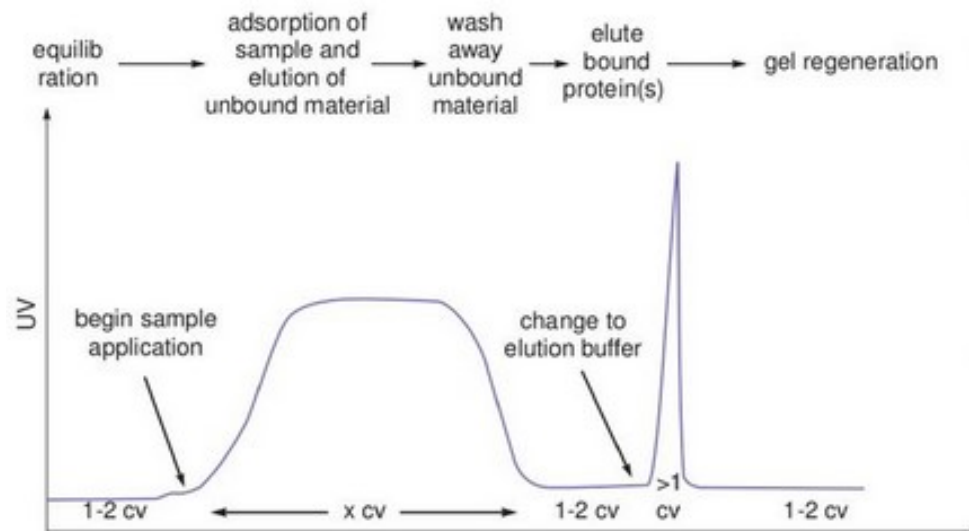
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:
AFFINITY

	GST	MBP	Strep™-tag II	His	FLAG™
Size	26 kDa	40 kDa	8 aa	6 aa	8 aa
Protein binding capacity	30 mg/ml	10 mg/ml	6 mg/ml	40 mg/ml	0.6 mg/ml
Purity	+ +	+ +	+ + +	+	+ + +
Increased solubility	Yes	Yes	No	No	No
Risk for interference with function	Yes	Yes	No	No	No

Immobilized GSH

Immobilized amylose

Stabilized Strep-Tactin Tetramer
Binding of twin-Strep-tagged Bait Protein
Non-Interacting Proteins
Immobilized Streptavidin/Strep-Tactin

PROTEIN
NTA
Acido nitrilotriacetico

Solid phase
Tag
Target protein
Anti tag-antibody
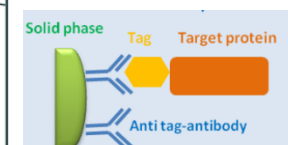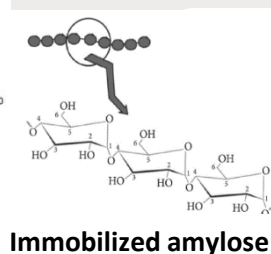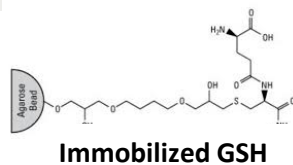
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

REVERSE AFFINITY
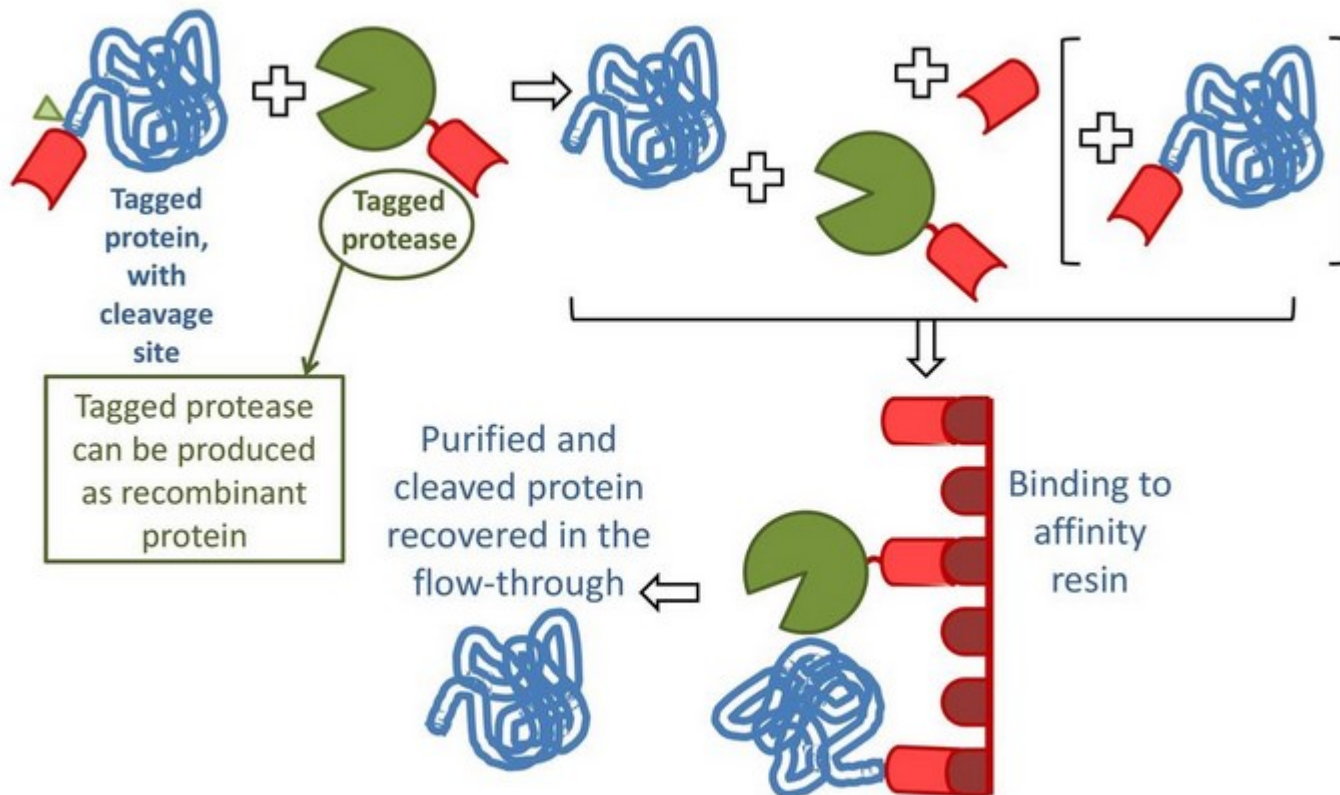
Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:

Method	Typical characteristics		Purification phase			Sample start conditions	Sample end conditions
	Resolution	Capacity	Capture	Intermediate	Polishing		
AC	+++ or ++	+++ or ++	+++	++	+	Various binding conditions	Specific elution conditions
IMAC	+++	++	+++	++	+	For purifying histidine-tagged proteins using Ni Sepharose columns:: 20-40 mM imidazole; pH > 7; 500 mM NaCl; no chelators Other proteins: low concentration of imidazole	High concentration of imidazole, pH > 7, 500 mM NaCl
GF	++	+	+		+++	Most conditions acceptable, limited sample volume	Buffer exchange possible, diluted sample
IEX	+++	+++	+++	+++	+++	Low ionic strength. pH depends on protein and IEX type	High ionic strength or pH changed
HIC	+++	++	++	+++	+++	High ionic strength, addition of salt required	Low ionic strength

Purificazione di proteine ricombinanti per studi strutturali
Metodi cromatografici:
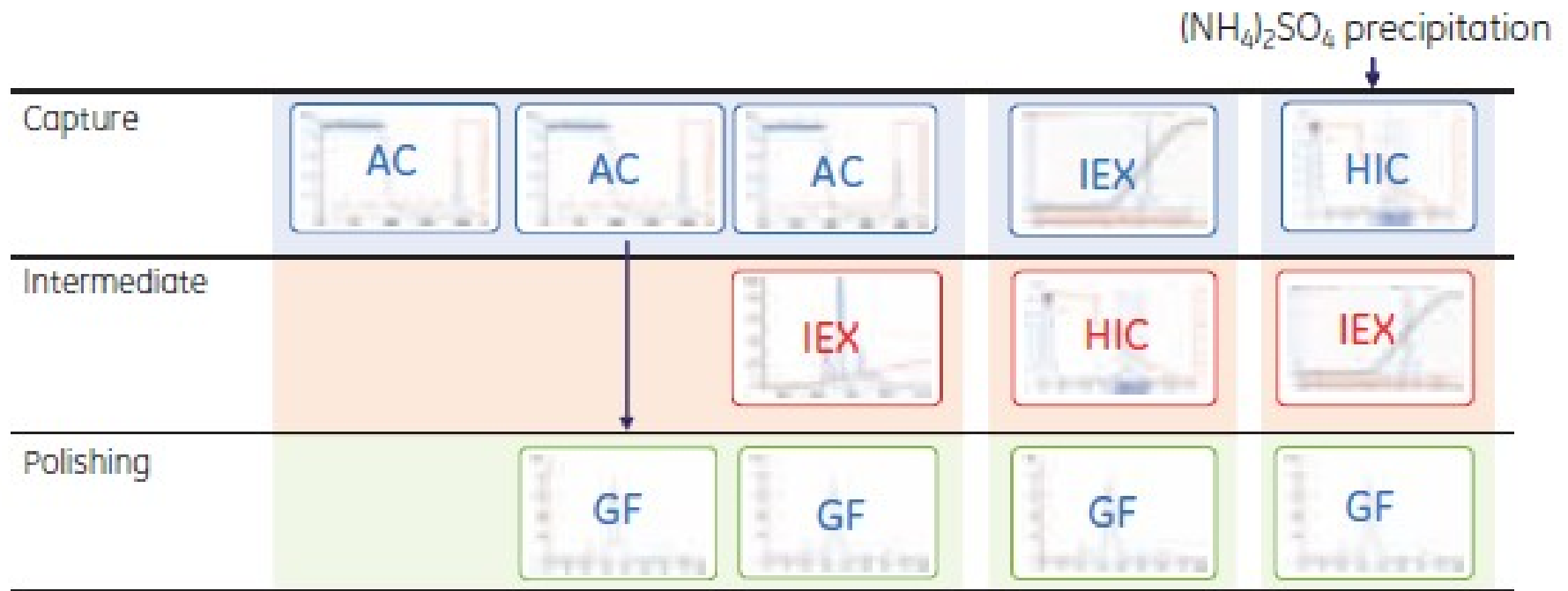
Espressione di proteine ricombinanti per studi strutturali
Espressione in forma insolubile e refolding:
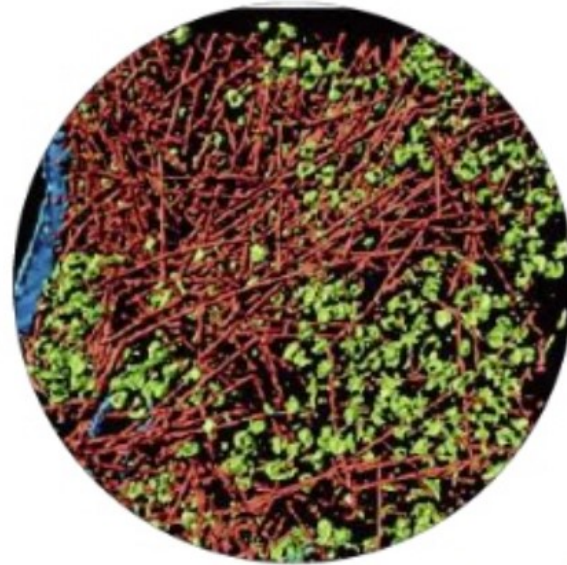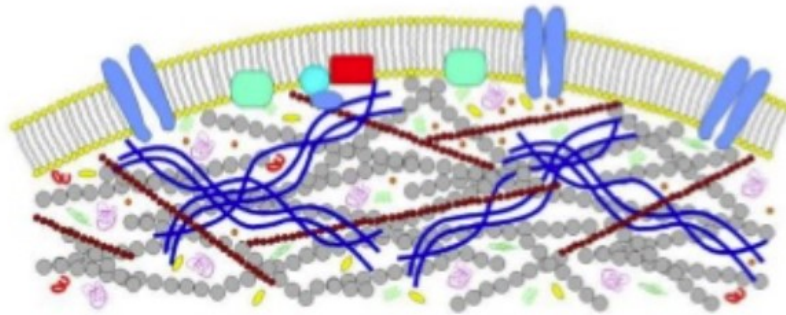
Folding in vivo: cytoplasm crowding and confinement

Espressione di proteine ricombinanti per studi strutturali
Espressione in forma insolubile e refolding:
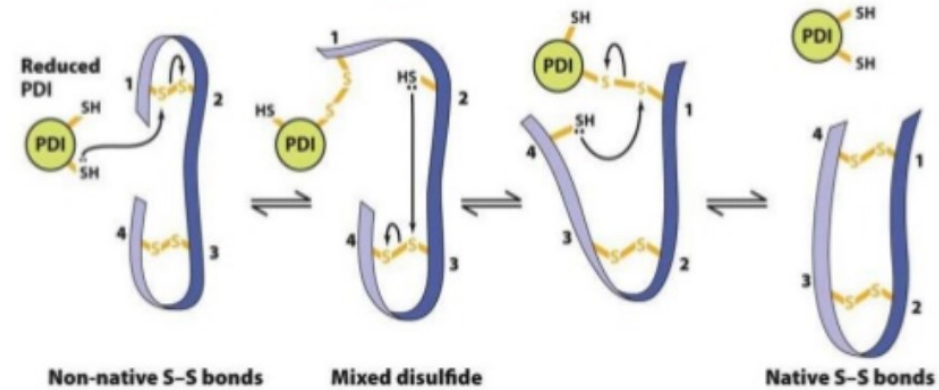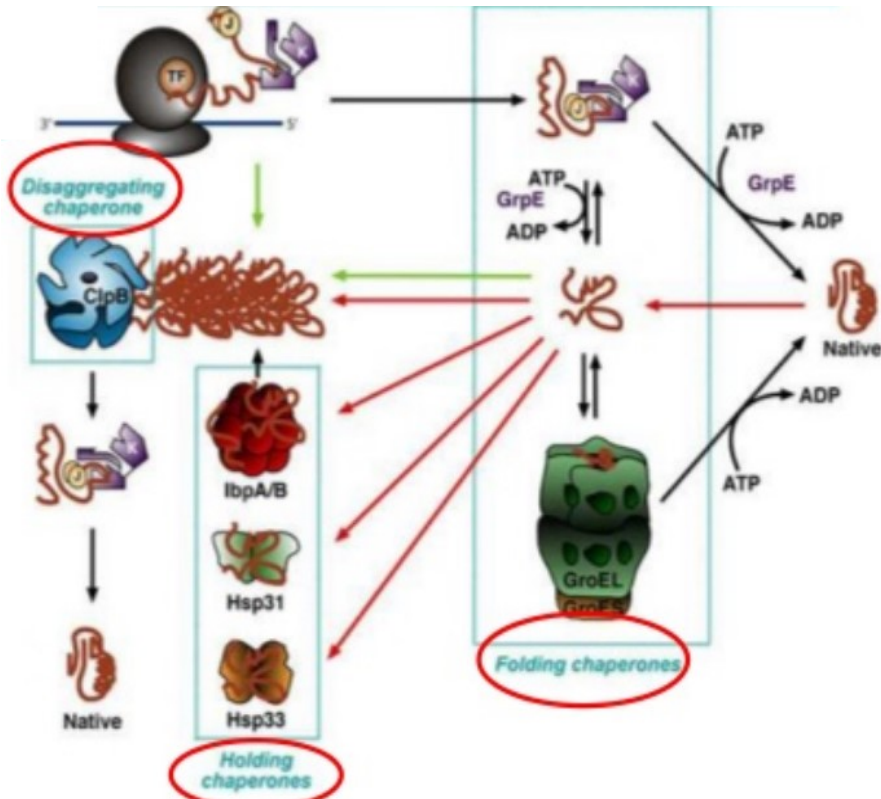
Folding in vivo: ruolo delle chaperonine/foldasi/Protein Disulfide Isomerase

Espressione di proteine ricombinanti per studi strutturali
Espressione in forma insolubile e refolding:

L'overespressione satura il sistema enzimatico:
aggregazione/accumulo di proteine misfolded in corpi di inclusione:

Vantaggi:

Protezione da proteasi citoplasmatiche
Elevate rese
Elevata purezza
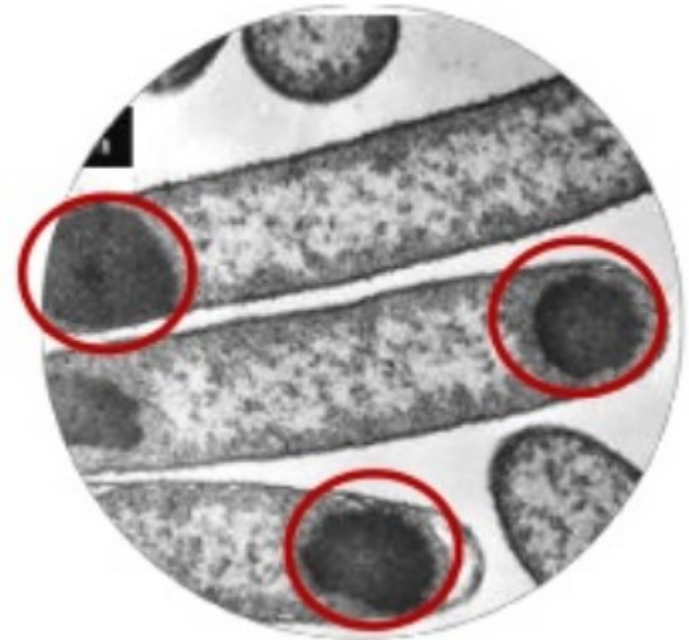
Espressione di proteine ricombinanti per studi strutturali
Espressione in forma insolubile e refolding:

Centrifugazione differenziale

SEC

Espressione di proteine ricombinanti per studi strutturali
Espressione in forma insolubile e refolding:

Correctly folded

Incorrectly folded
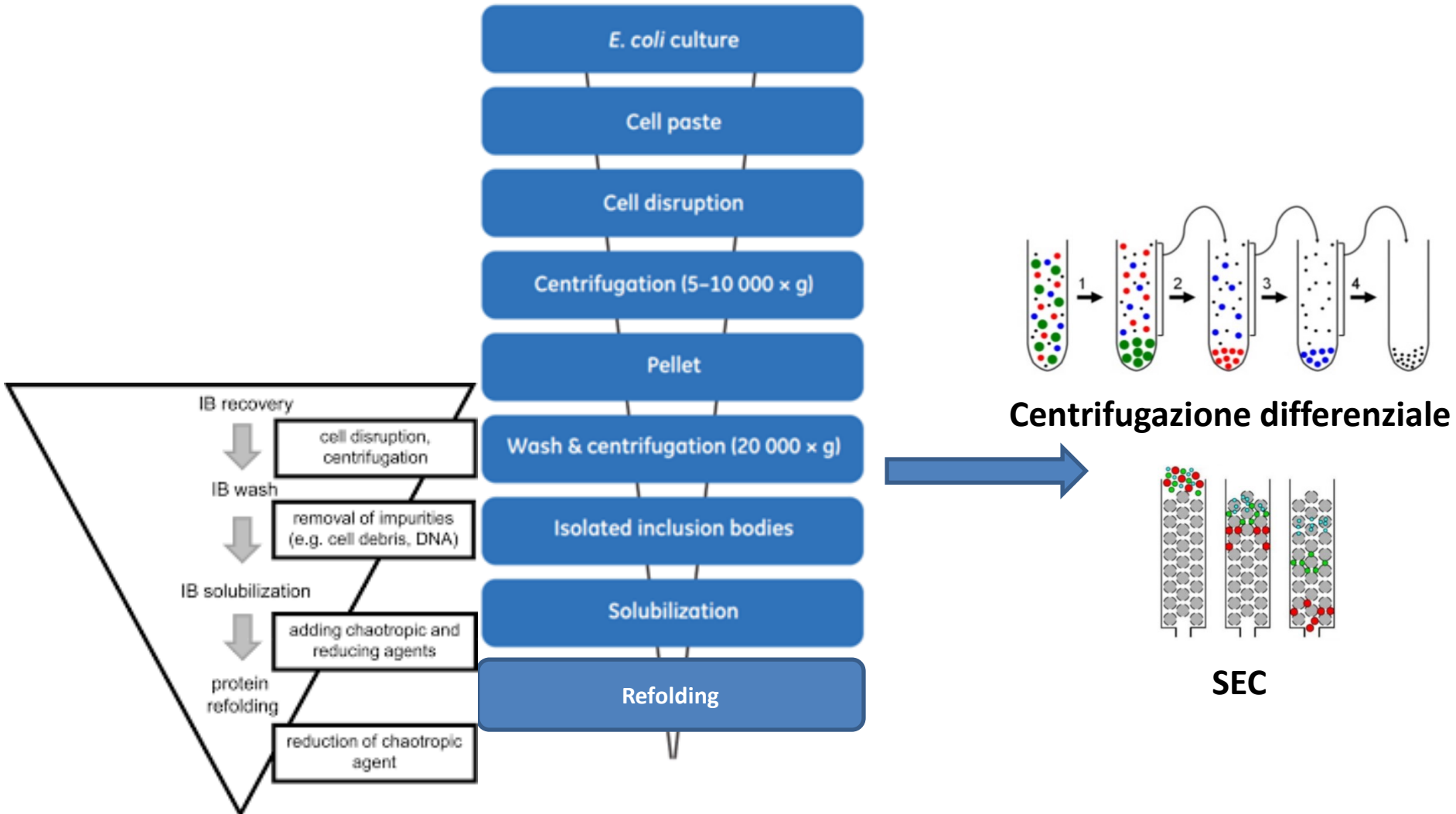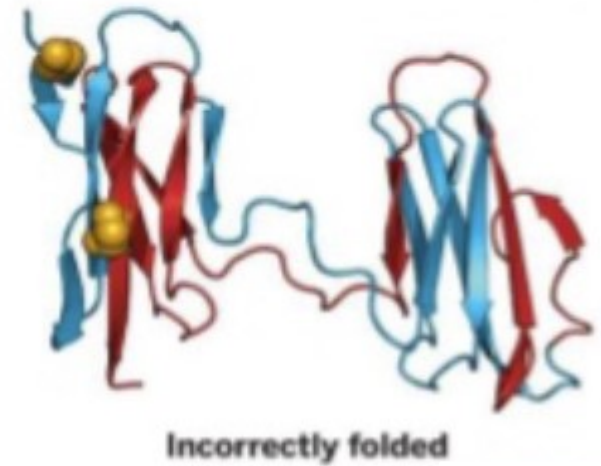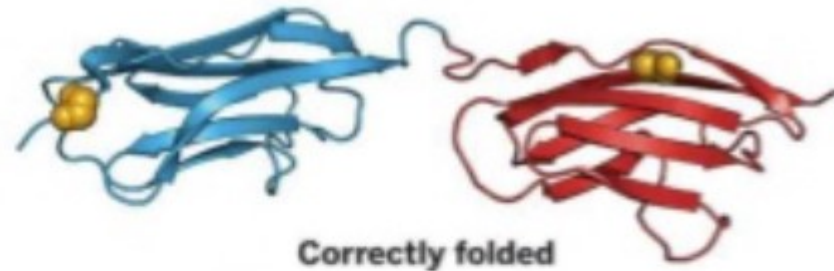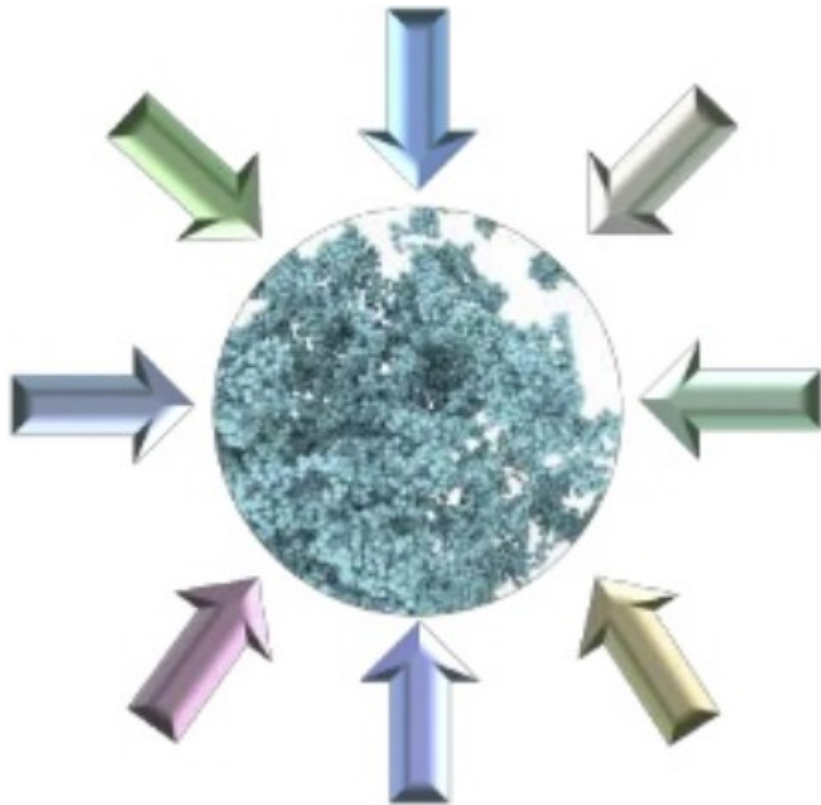
Misfolding ⟷ **Refolding** ⟷ **Aggregazione**

Espressione di proteine ricombinanti per studi strutturali
Espressione in forma insolubile e refolding:
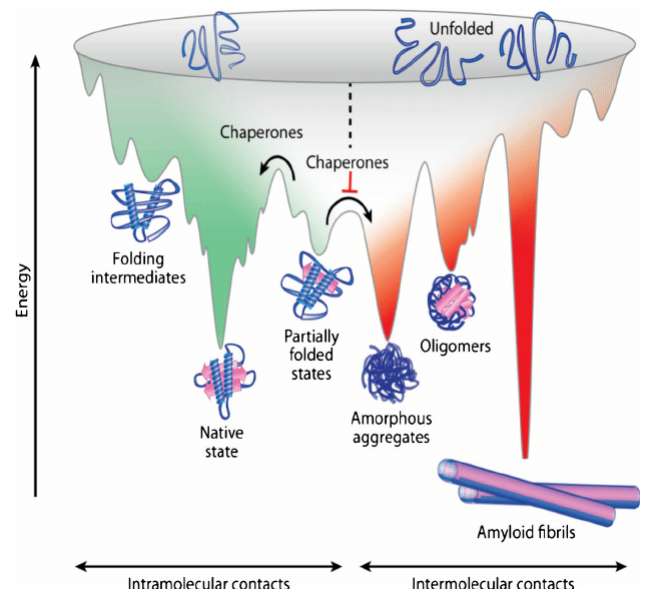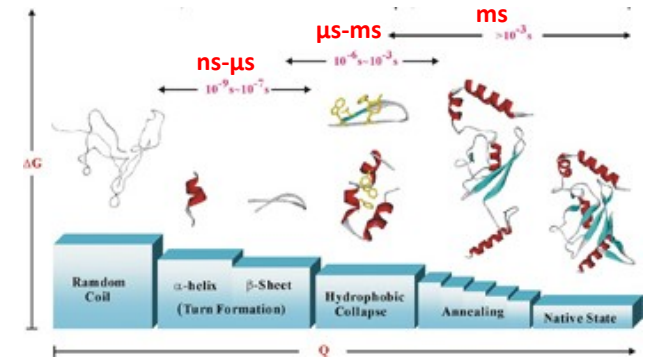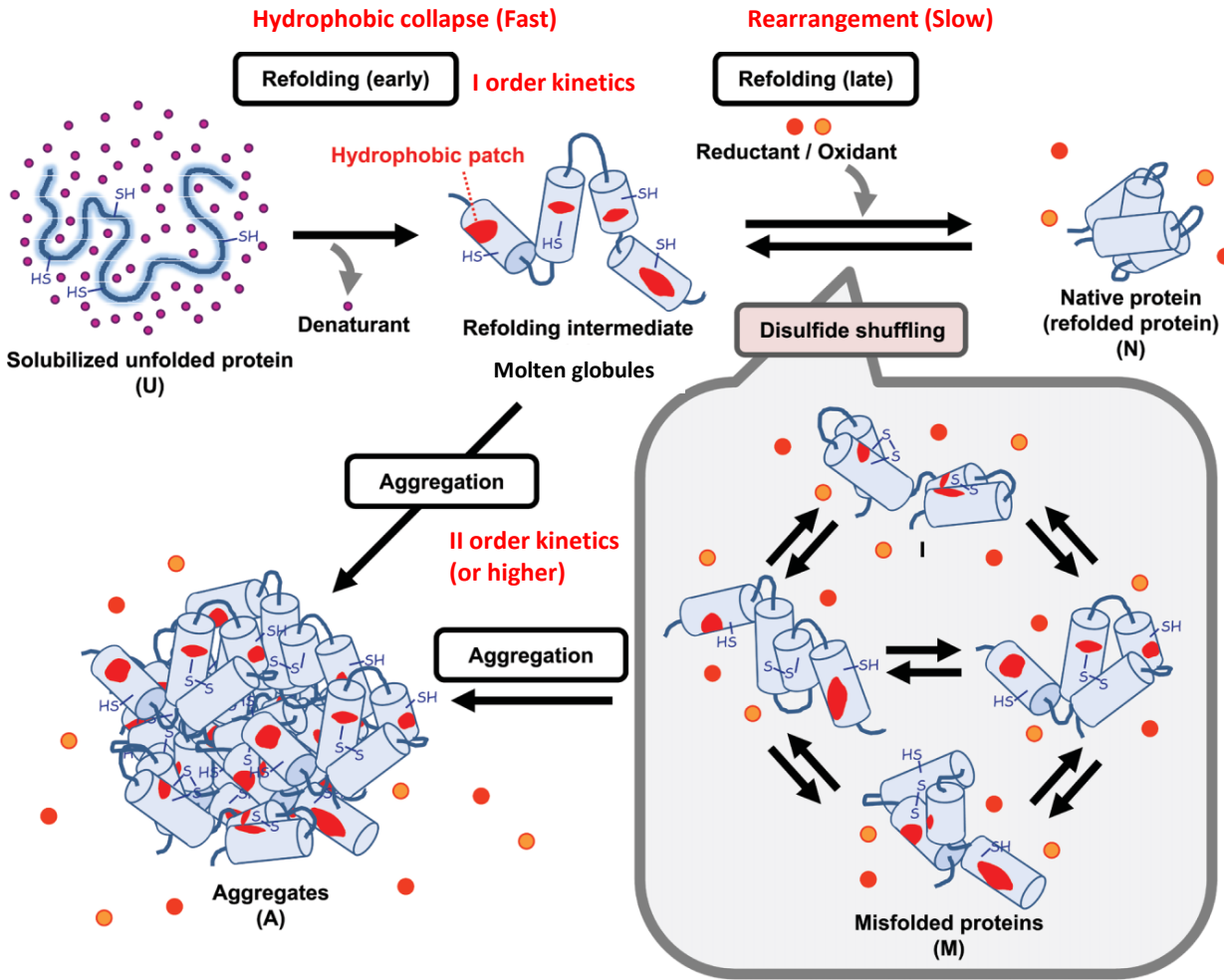
Driving forces: interazioni idrofobiche> WdW> Pro isomerization > ponti di solfuro> legami H

Da ottimizzare vari parametri (T, t, pH, forza ionica)
Impiego di additivi

Additive	Concentration	Effect
CHAPS	30 mM	Detergent
EDTA	20 mM	Chelate
Glycerol	10–50%	Stabilizer
Guanidine HCl	0.1–1 M	Chaotroph
L-arginine	0.4–0.5 M	Stabilizer
Laroylsarcosin	Up to 4 M	Detergent
MgCl$_2$/CaCl$_2$	2–10 mM	Cation
NaCl/Ammonium sulfate	0.2–0.5 M	Salt
Non-detergent sulfo betain	Up to 1 M	Solublizer
PEG 3350 and cyclodextrin	Up to 0.5% W/V	Osmolyte
SDS	0.1%	Detergent
Sodium citrate/sulfate	0.2–0.5 M	Salt
Sucrose/glucose	Up to 0.75 M	Stabilizer
TMAO	1–3 M	Osmolyte
Tris	0.4–1 M	Buffer
Triton X-100	0.1–1%	Detergent
Tween-80	0.01%	Detergent
Glycine	Up to 1 M	Osmolyte
Proline	Up to 1 M	Osmolyte
Urea	0.1–2 M	Chaotroph
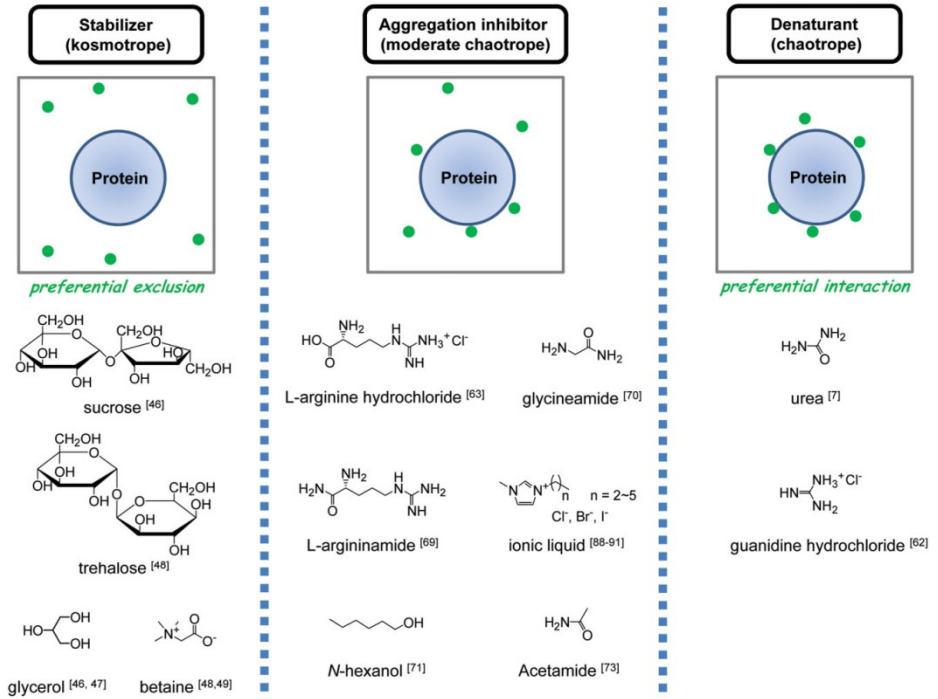
Artificial chaperones
Detergents: Triton X-100, CHAPS, Sarkosyl, SDS and CTAB
Detergents with cycloamylose or cyclodextran
Non-detergent zwitterionic agents such as Non-detergent sulfobetaine
(NDSB), substituted pyridines and pyrroles,
and substituted amino-cyclohexanes

Espressione di proteine ricombinanti per studi strutturali
Metodi di refolding: il tampone di refolding

Impiego di addittivi proteici

Chaperone mediated (soluble or immobilized and reusable system format)
ATP dependent: GroEL-GroES, DnaK, DnaJ, GrpE minichaperones
ATP independent: peptidyl-prolyl isomerase
                                    DsbA, Trx and GST

Antibody mediated
specific for conformational epitopes

Impiego di agenti redox

Reducing/oxidizing reagents: BME, BMC, TCEP, DTT, DTE,
    cysteine/cystine, glutathione, cysteamine/cystamine, and EDTA

Espressione di proteine ricombinanti per studi strutturali
Metodi di refolding in soluzione:
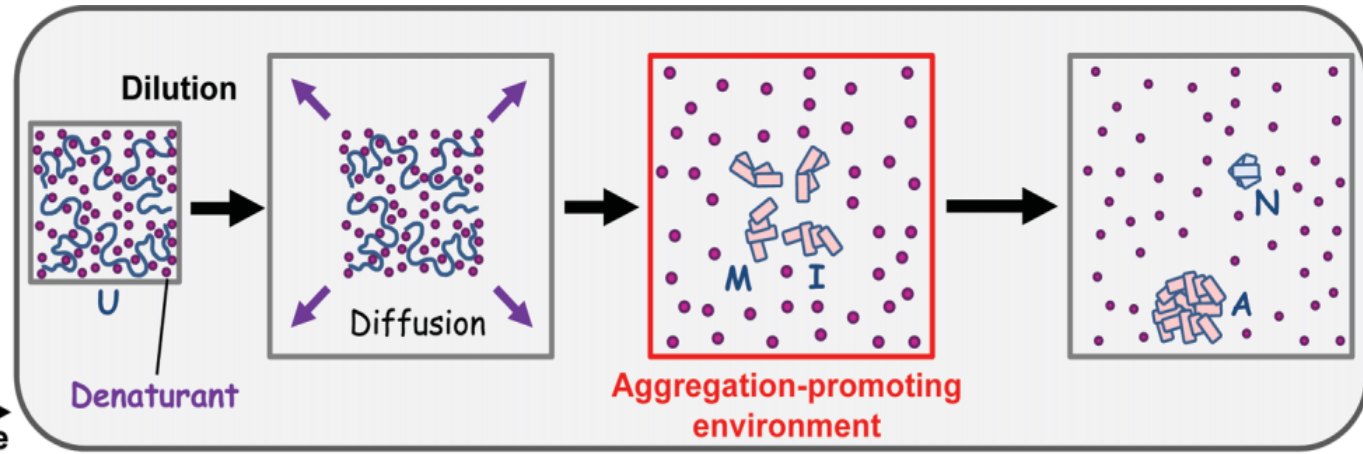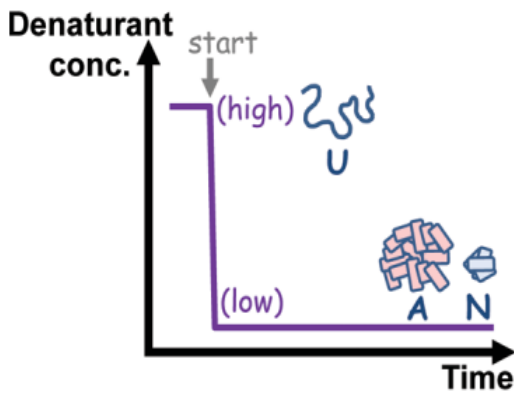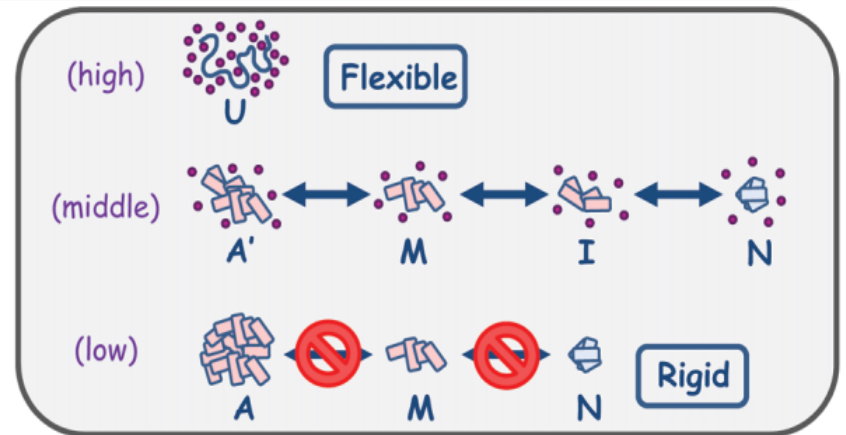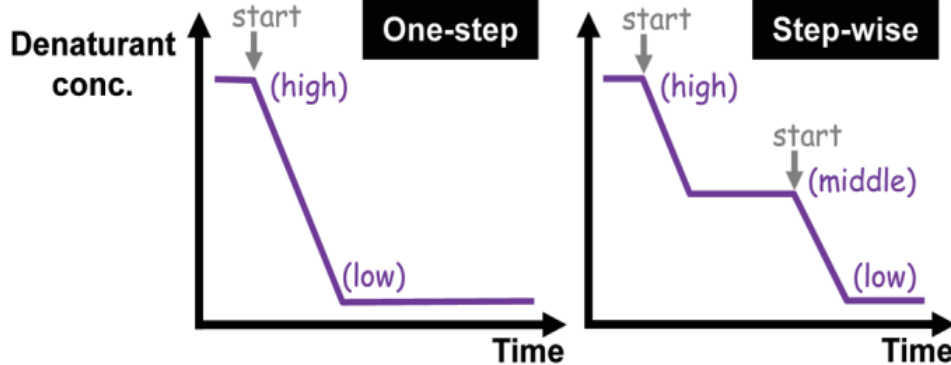
Espressione di proteine ricombinanti per studi strutturali
Metodi di refolding:

C) REFOLDING SU COLONNA
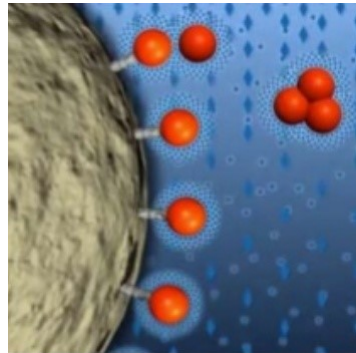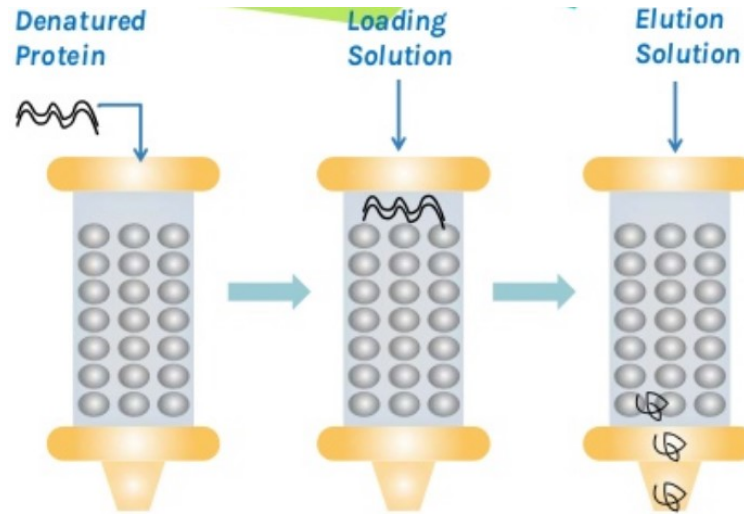
**Adsorbimento/Immobilizzazione sulla matrice inibiscono interazioni aspecifiche →
riducono l'aggregazione**

Espressione di proteine ricombinanti per studi strutturali
Refolding su colonna:

C. Solid phase method

SEC-based method

U   M   I   A   N   Denaturant
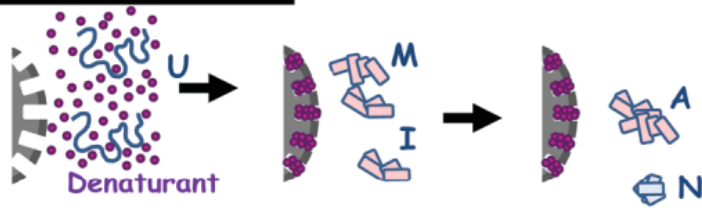
Immobilization method

Ligand   Tag   U   Eluent   I   N
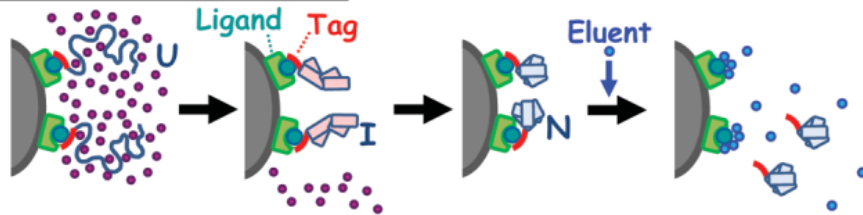
IMAC

Adsorption method

Eluent   I   A   N
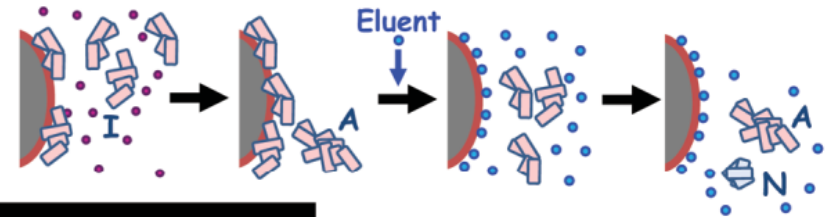
Zeolite-based method

U   Eluent   A   N
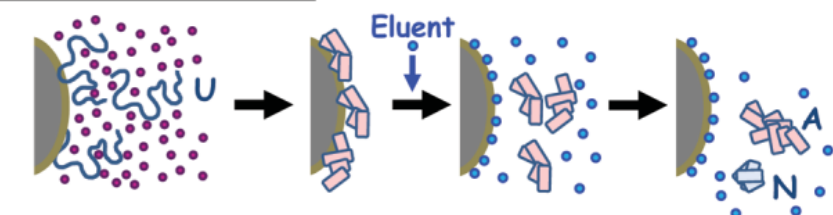
Espressione di proteine ricombinanti per studi strutturali
Metodi di refolding:

ADSORBIMENTO

A) IEX
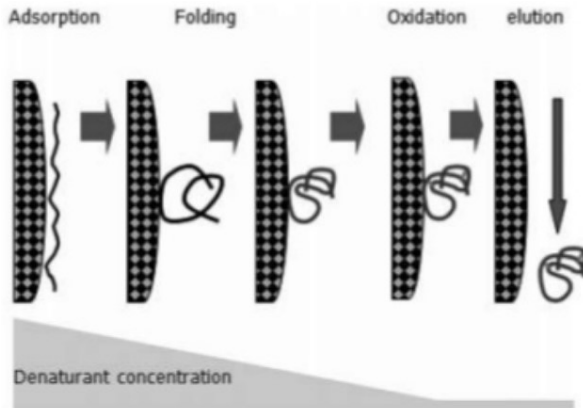
B) HIC

Gradiente duplice:

1) pH più elevato del pI
(riduce l'aggregazione)
2) pH vicino al pI (compatibile
con il folding nativo)
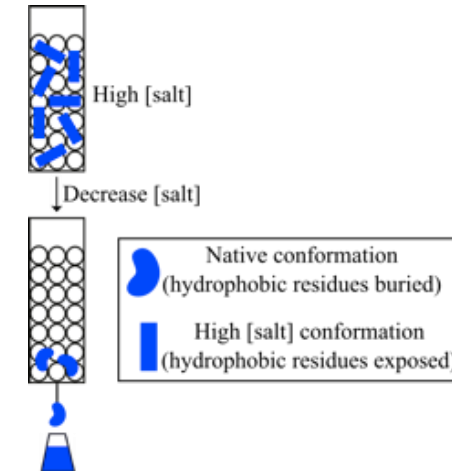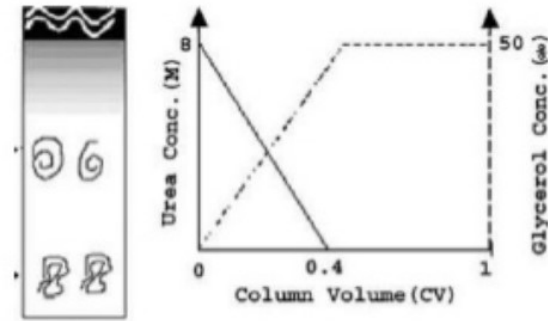
Produzione di proteine ricombinanti per studi strutturali
Cristallizzazione: considerazioni energetiche
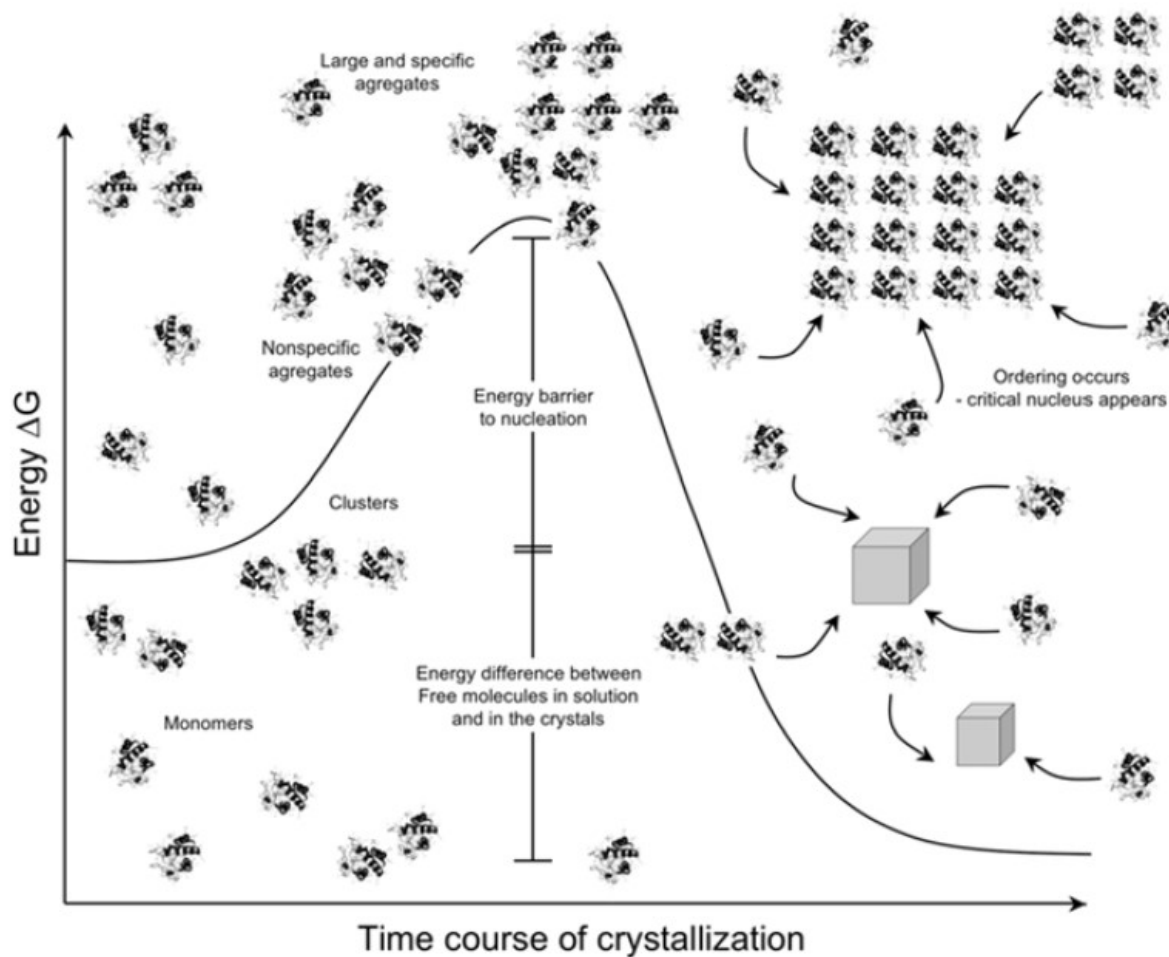
$$\Delta G_{cryst} = \Delta H_{prot} - T\Delta S_{prot} - T\Delta S_{solv}$$

Produzione di proteine ricombinanti per studi strutturali
Cristallizzazione: considerazioni energetiche

autoassemblaggio in cristalli:

- formazione di numero di **contatti relativamente piccolo/interazioni deboli**:

(hydrogen bonds,salt bridges,p-p,s-p,charge-p interactions, *etc)*
ΔH_c negativi o poco significativi **(–17/0 kcal/mol)**

- perdita significativa dei gradi di libertà traslazionale/rotazionale

conformazionale dei residui coinvolti

$\Delta S_{protein}$ 30/100 kJ / mol $\rightarrow$ **-T$\Delta S_{protein}$ =+7,5/25 kcal / mol (a T$_{amb}$)**

contributo destabilizzante positivo alla ΔG_c (perdita di gradi di libertà della proteina)

- compensato dal rilascio di solvente attraverso residui di superficie sia idrofobici che polari durante l'impacchettamento (guadagno entropico del solvente)

$\rightarrow$ **-TΔS_{sov} =+25/75 kcal / mol (a T$_{amb}$)** corrisponde al rilascio di **circa 5-30 molecole** d'acqua ordinate dalla superficie

$\Delta G_c = \Delta H_c - T(\Delta S_{protein} + \Delta S_{solvent})$

Piccoli cambiamenti $\Delta S_{protein}$ / $\Delta S_{solvent}$ $\rightarrow$ effetti drammatici (SER: mutazione di residui superficiali con elevata libertà entropica (Lys, Glu, Gln in Ala)

Produzione di proteine ricombinanti per studi strutturali
Cristallizzazione: parametri chimico-fisici critici

Database = 821 proteine funzionalmente monomeriche

Acta Crystallographica Section D
BIOLOGICAL CRYSTALLOGRAPHY
journals.iucr.org/d
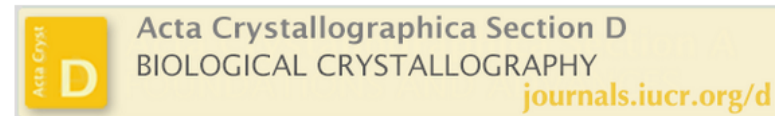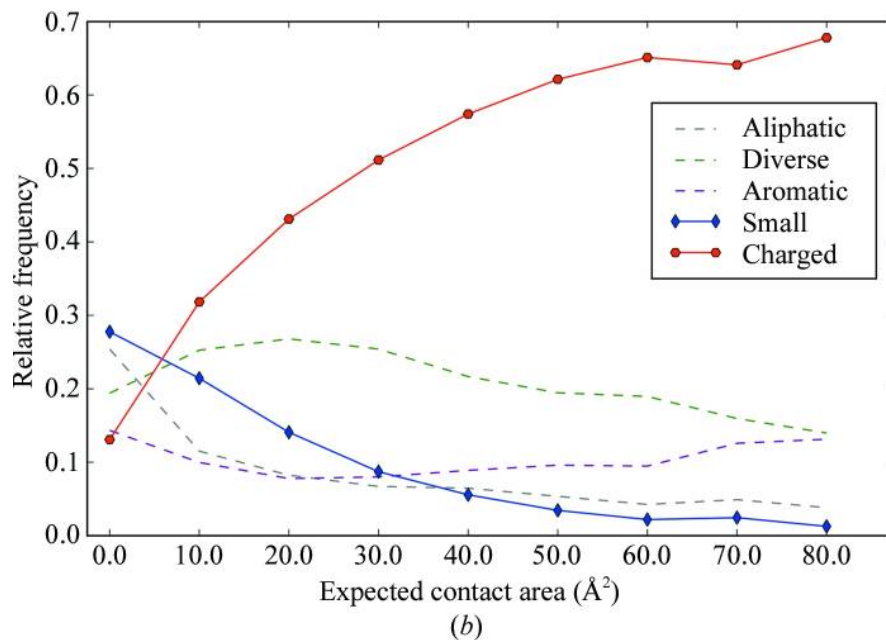
The role of entropy and polarity in intermolecular contacts in protein crystals

Marcin Cieślik and Zygmunt S. Derewenda

superficie accessibile al solvente dei residui

entropia delle catene laterali

Produzione di proteine ricombinanti per studi strutturali
Cristallizzazione: parametri chimico-fisici critici

Database = 679 proteine espresse (179 cristallizzano)
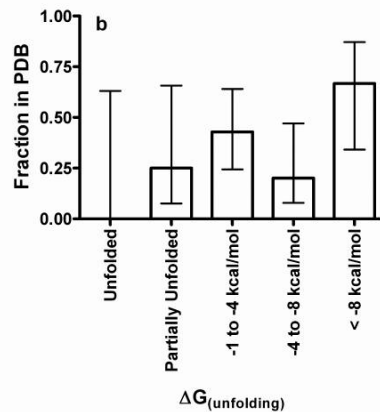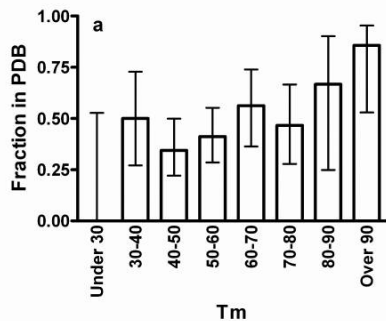
stabilità
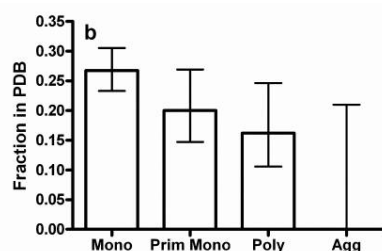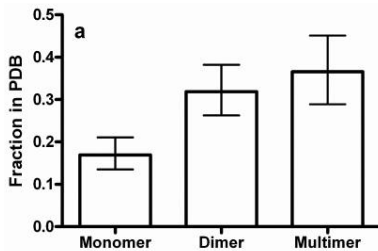
proprietà idrodinamiche

epitopi superficiali ordinati (G) → **interazioni stereospecifiche nell' impacchettamento**

Produzione di proteine ricombinanti per studi strutturali
Cristallizzazione: parametri critici
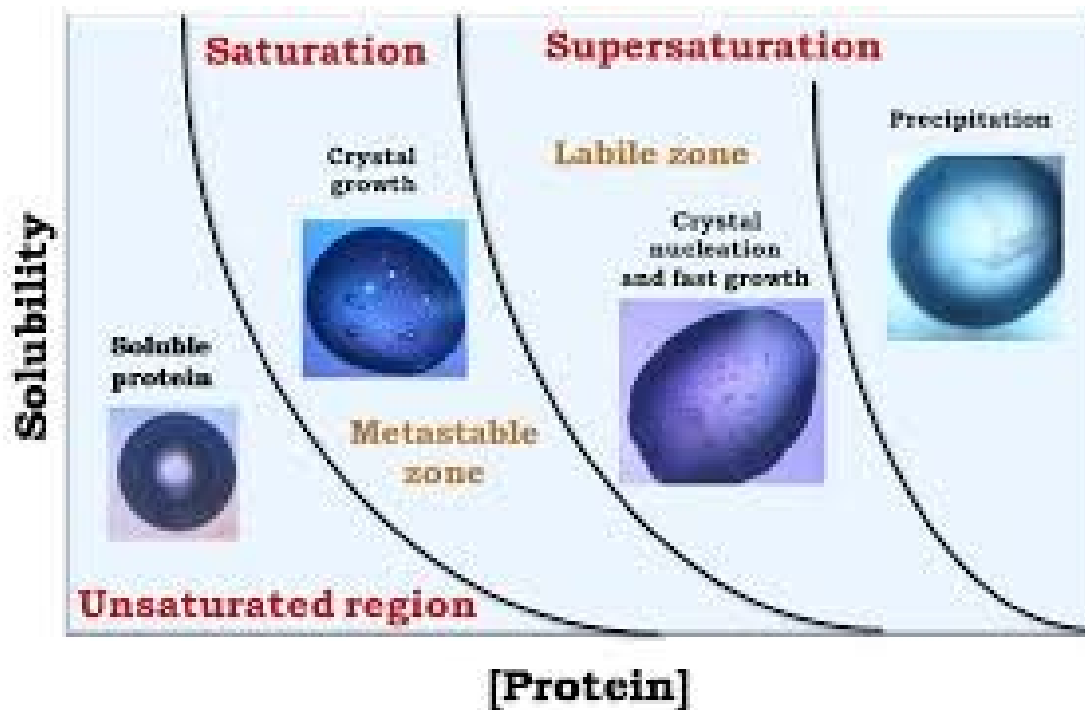
regioni con solvente strutturato
Gravy (Grand average hydrophobicity) **index**

Flessibilità/regioni disordinate
pI

1) FAb

Crystal structure of full-length KcsA in its closed conformation

Serdar Uysal[a,b], Valeria Vásquez[a,b], Valentina Tereshko[a,b], Kaori Esaki[a], Frederic A. Fellouse[c], Sachdev S. Sidhu[c], Shohei Koide[a], Eduardo Perozo[a,b,1], and Anthony Kossiakoff[a,b,1]

[a]Department of Biochemistry and Molecular Biology, and [b]Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637; and [c]Department of Protein Engineering, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080

2) **fusion partners (ex: MBP)**

Structure of a signal transduction regulator, RACK1, from *Arabidopsis thaliana*

HEMAYET ULLAH,[1] ERICA LOUISE SCAPPINI,[2] ANDREA FLORENCE MOON,[3] LATANYA VERONICA WILLIAMS,[1,4] DAVID LEE ARMSTRONG,[2] AND LARS CHRISTIAN PEDERSEN[3]

[1]Department of Biology, Howard University, Washington, DC 20059, USA
[2]Laboratory of Neurobiology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA
[3]Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Solubilità

3) **direct surface engeneering**

| Amino acid at position 76	Solubility (mg/ml)[a]
Asp	43
Arg	42
Glu	42
Ser	39
Lys	31
Gly	27
Ala	27
His	24
Asn	21
Thr	20
Gln	20
Pro	15
Cys	12
Met	11
Val	10
Leu	9.3
Ile	8.2
Tyr	5.6
Phe	4.4
Trp	3.6
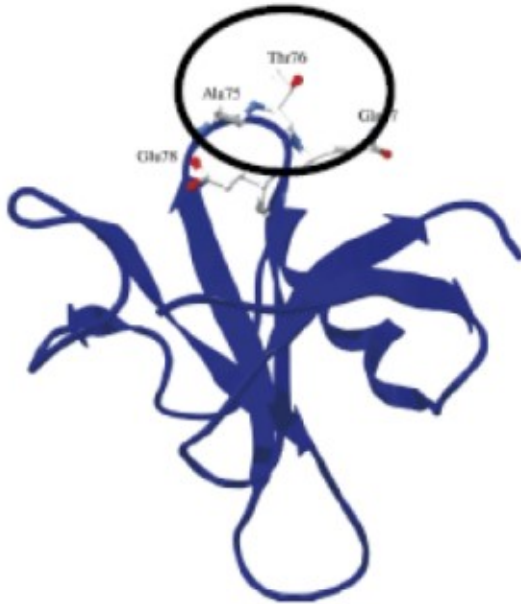
[a] The error in these measurements is ±10%.

Amino Acid Contribution to Protein Solubility: Asp, Glu, and Ser Contribute more Favorably than the other Hydrophilic Amino Acids in RNase Sa

Saul R. Trevino [1], J. Martin Scholtz [1, 2] ✉, C. Nick Pace [1, 2] ✉

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Stabilità/Solubilità

Ottimizzazione delle curve di denaturazione
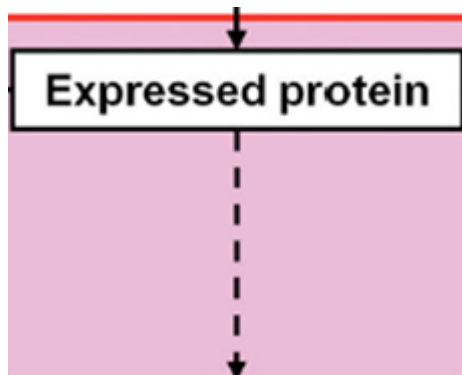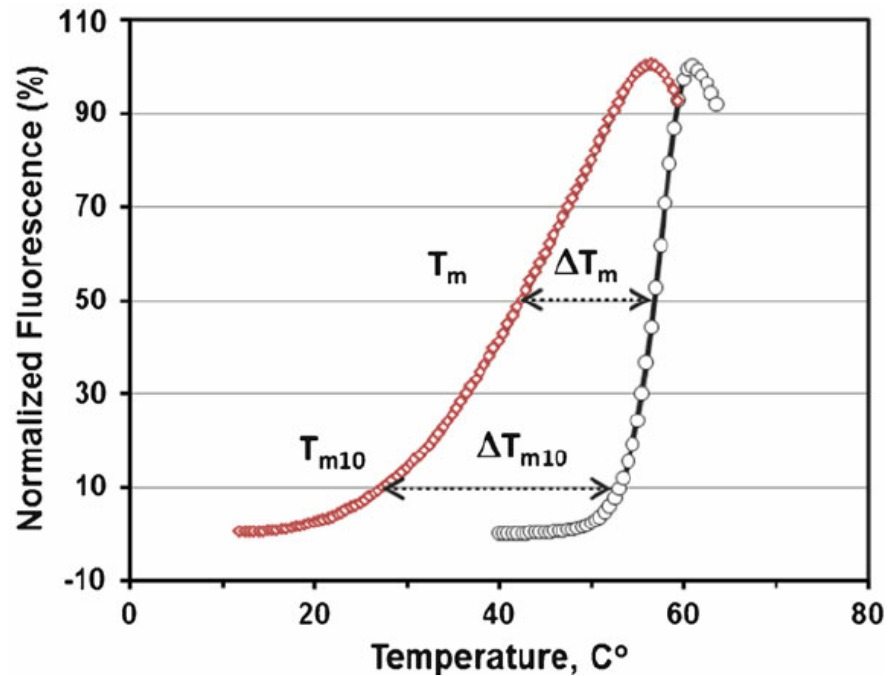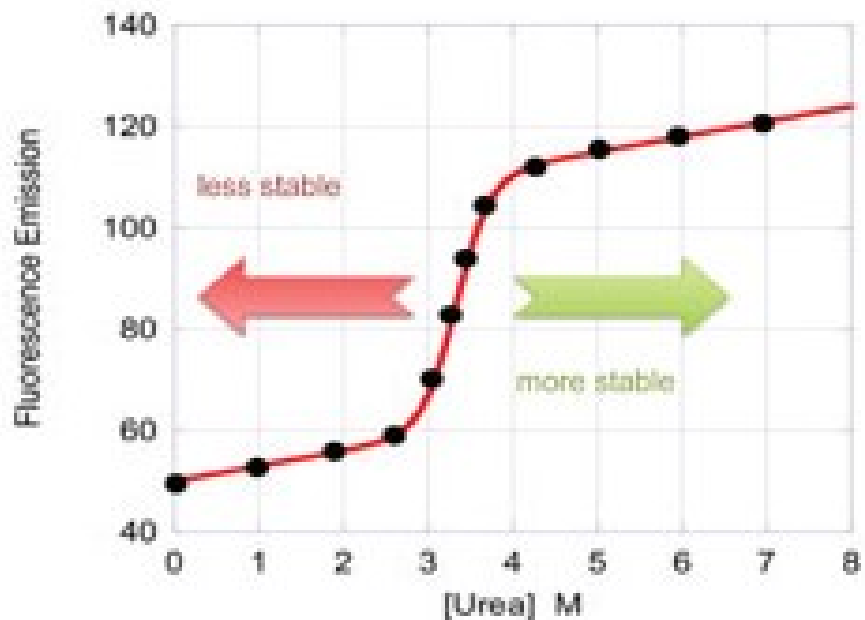
screening di formulazioni

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Stabilità/Solubilità

•**Biologia molecolare:**
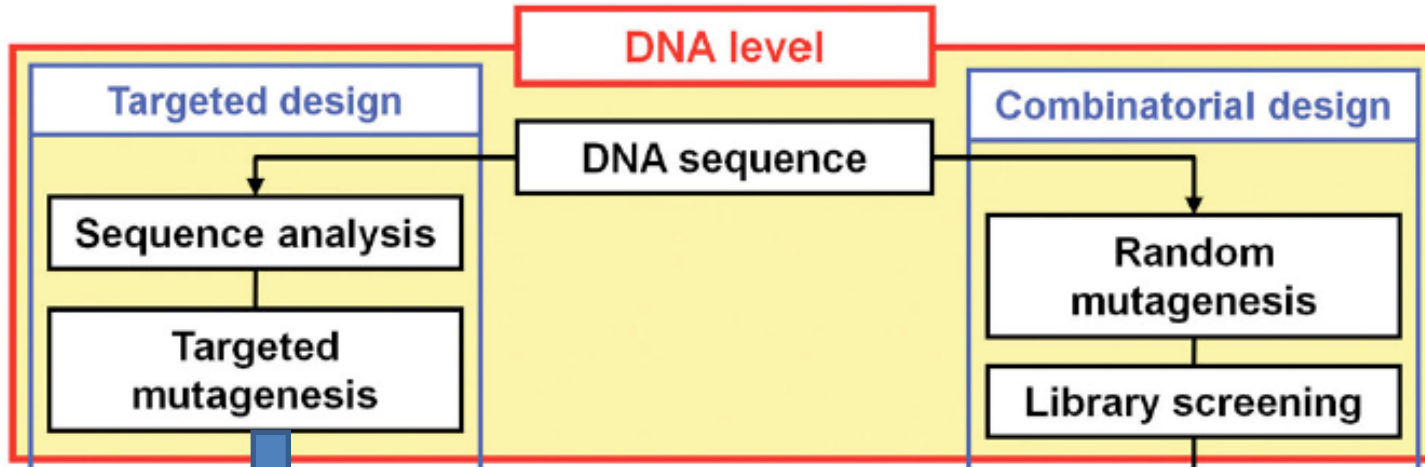
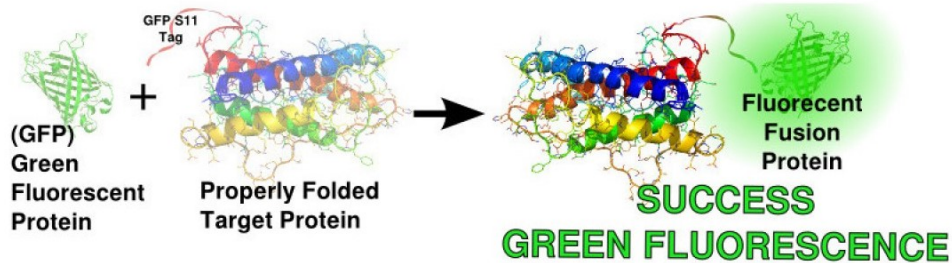**Mutagenesi sito specifica:**

1. QuikChange
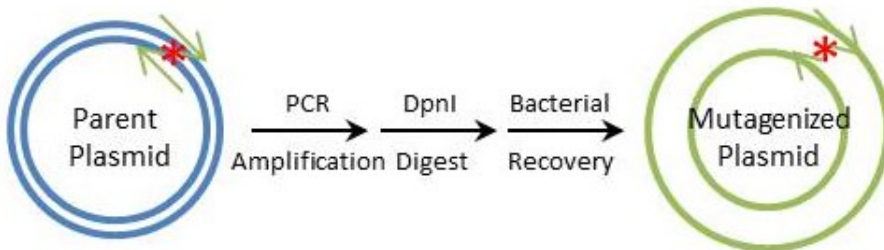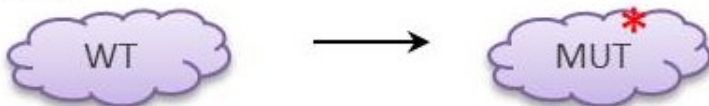
2. overlap- assembly PCR
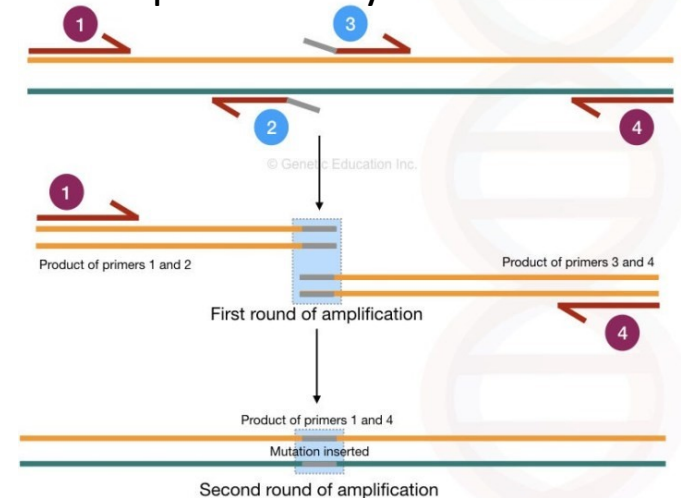
Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Stabilità/Solubilità

Directed evolution
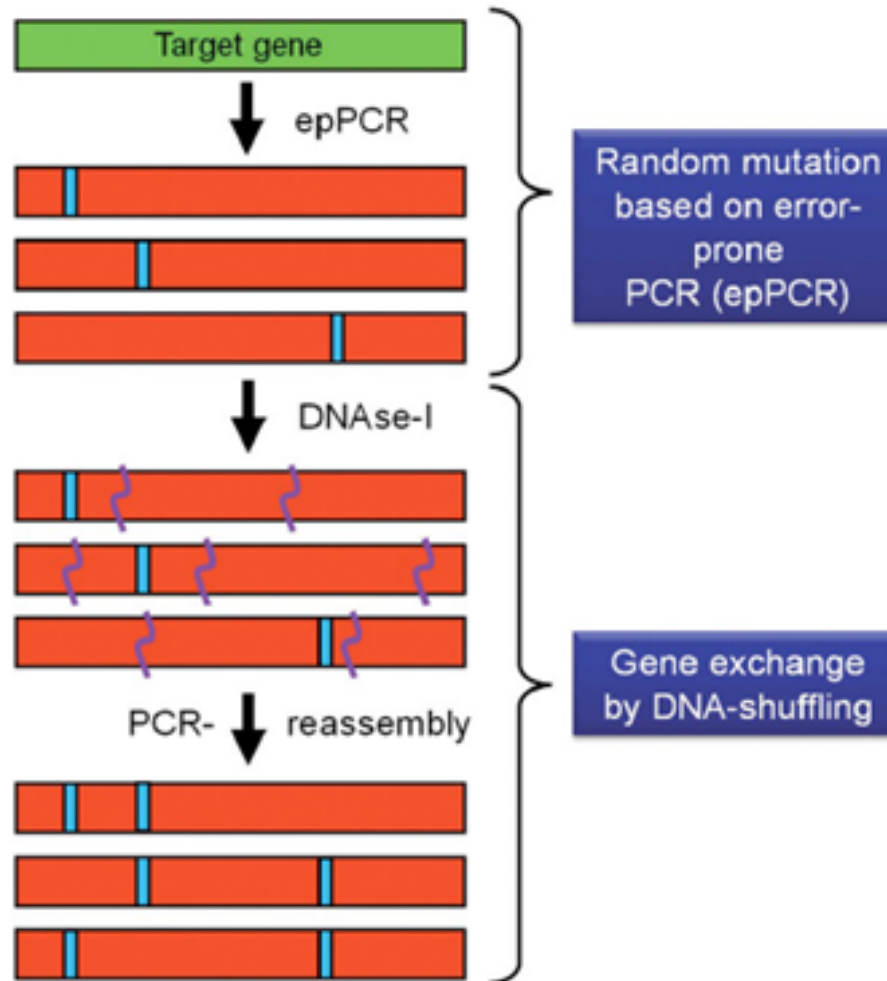
© Garland Science 2010

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Stabilità/Solubilità

Read out: fusione con GFP
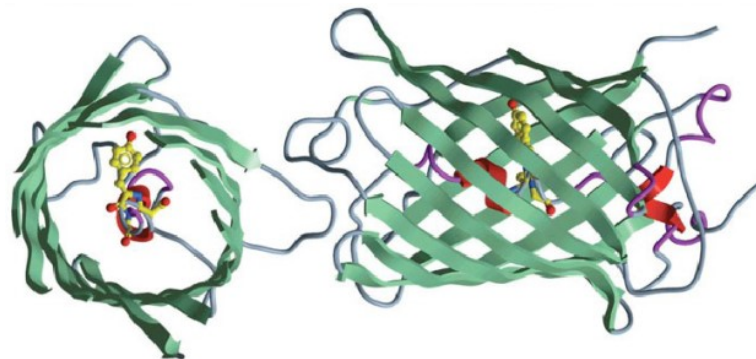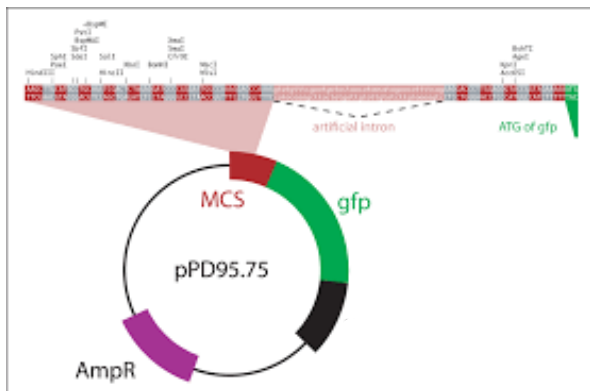
S65-Y66-G67: ciclizzazione, deidratazione e ossidazione→
p-idrossibenzilidene-*imidazolinone*

**Bacterial Fold-N-Glow™
Split GFP S11 Plasmid**
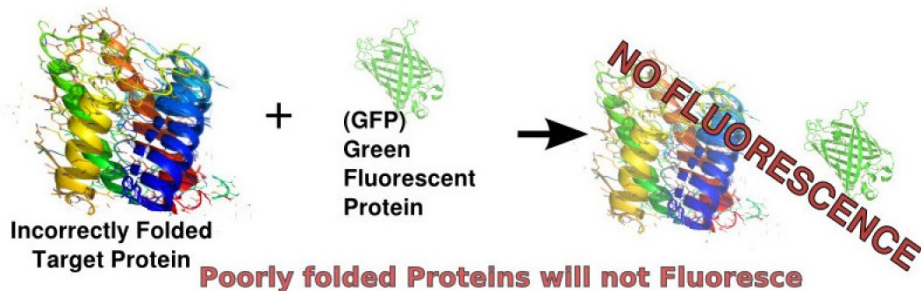
Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Stabilità/Solubilità

Read out: fusione con GFP

© Garland Science 2010
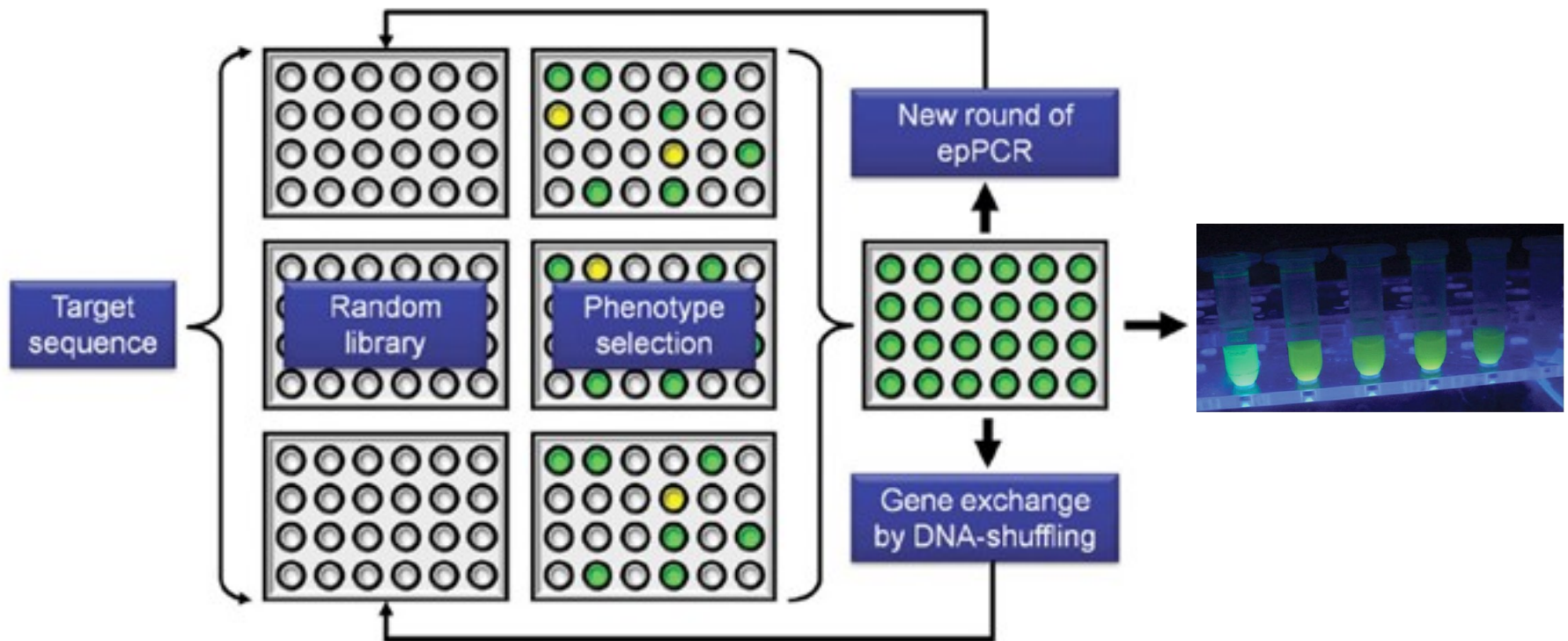
Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Abbassamento della barriera entropica

• **Eliminare modifiche post-traduzionali:**   **Glicosilazioni:**

1. **Mutagenesi**
2. **Inibitori**
3. **Glicosidasi**
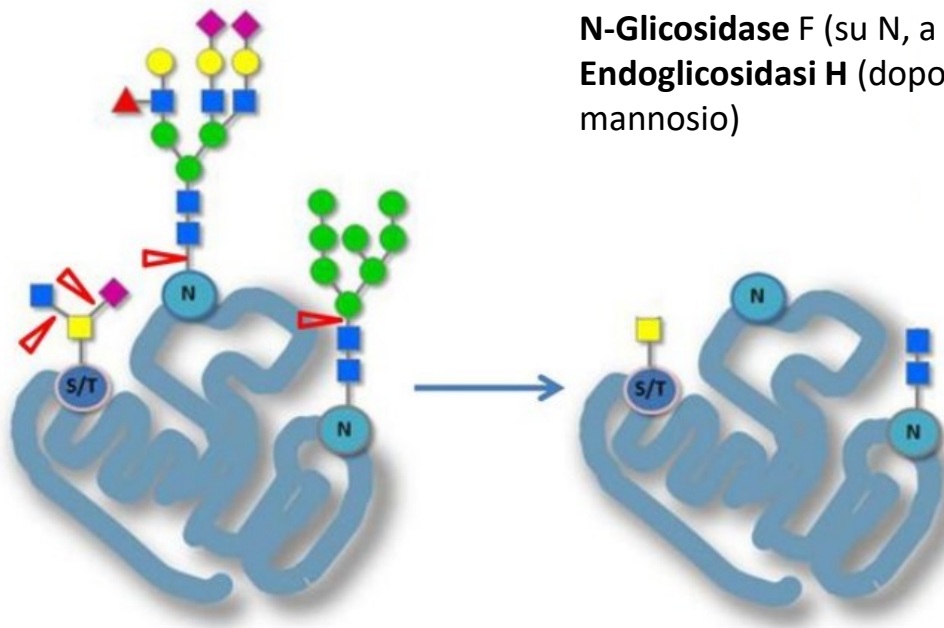
N-Glicosidase F (su N, a monte della prima N-acetilglucosamina)
Endoglicosidasi H (dopo la prima N-acetilglucosamina in glicani ricchi di mannosio)

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Abbassamento della barriera entropica

•**Eliminare flessibilità/tags (vedi analisi bioinformatica): biologia molecolare/proteolisi**
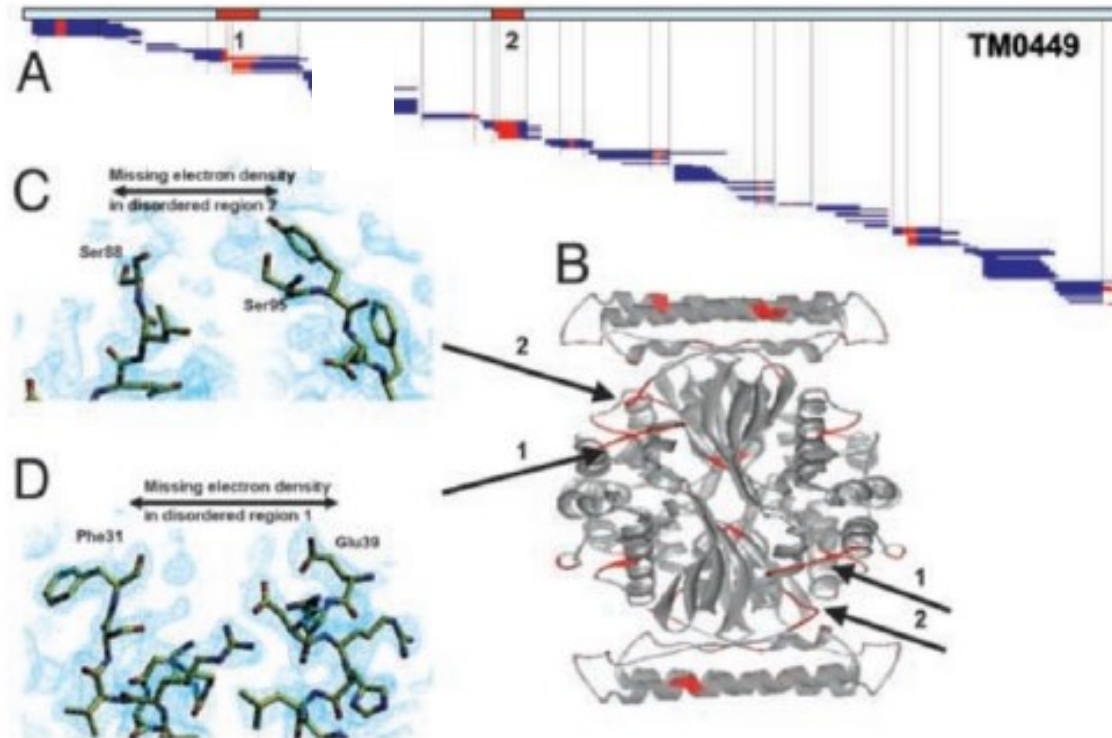
Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS

Dennis Pantazatos*, Jack S. Kim*, Heath E. Klock†, Raymond C. Stevens‡, Ian A. Wilson‡, Scott A. Lesley†, and Virgil L. Woods, Jr.*§¶

*Department of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA, 92093; †Joint Center for Structural Genomics, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; and ‡Joint Center for Structural Genomics, Genomics Institute of the Novartis

Produzione di proteine ricombinanti per studi strutturali
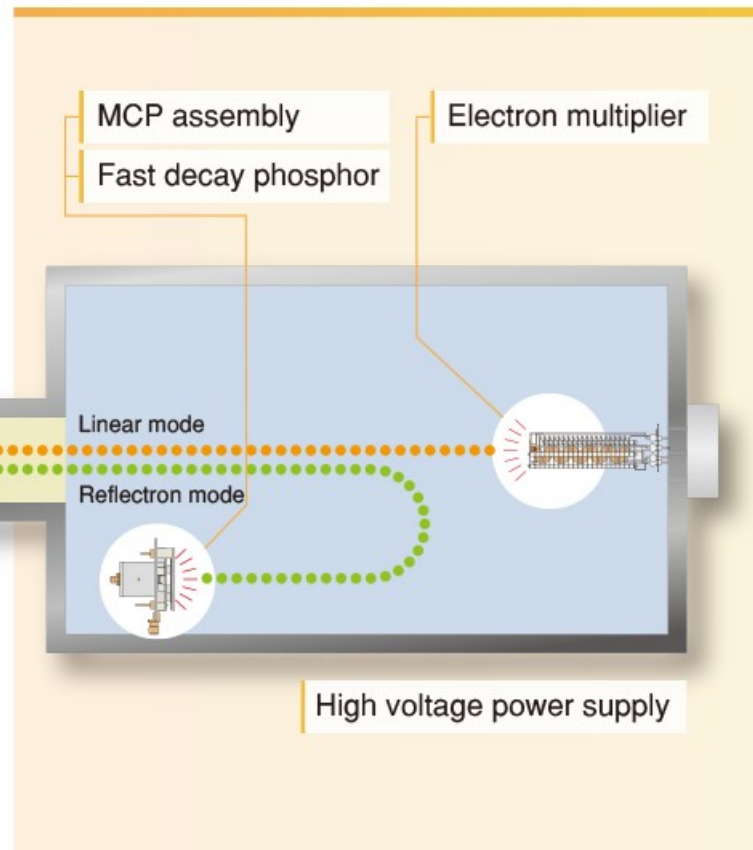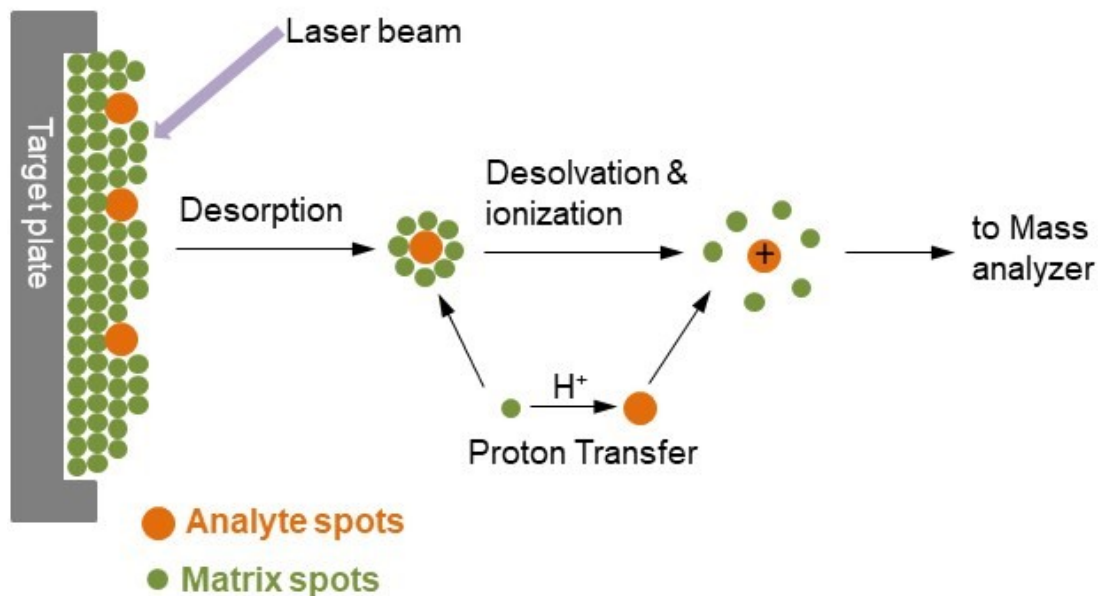## Spettrometria di massa
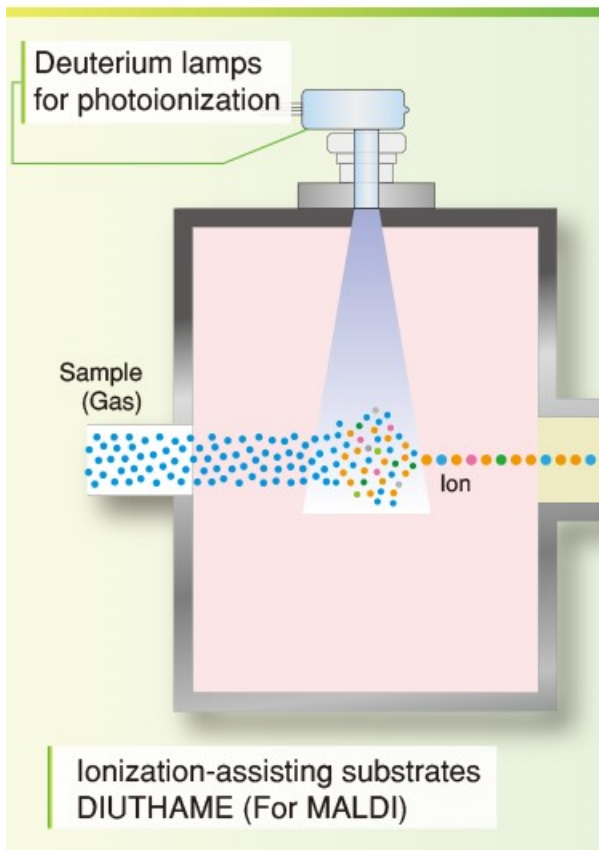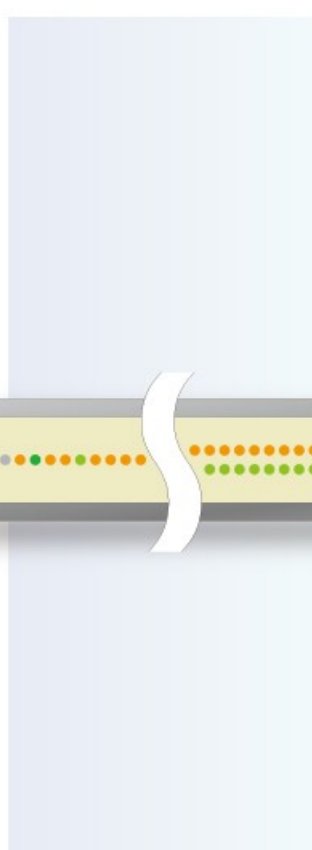
Produzione di proteine ricombinanti per studi strutturali
Spettrometria di massa: ionizzazione

MALDI (Matrix Assisted Laser Desorbtion Ionization)

Matrici
CHCA (alpha-Cyano-4-hydroxycinnamic acid)   MW < 10kDa
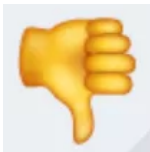SA (sinapinic acid)   > 10kDa
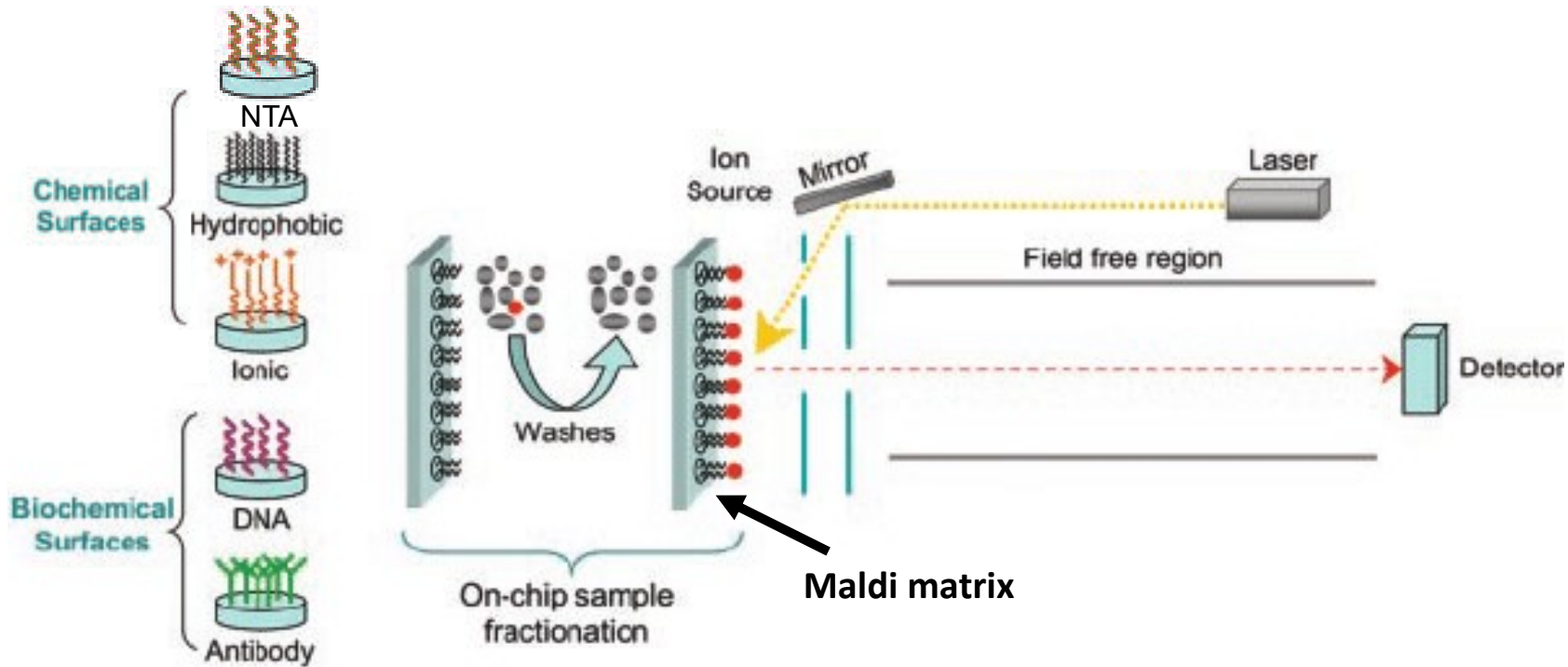Super DHS   glicoproteine

👍 Semplice
MW alti

👎 interferenza fra 50-500 m/z (non quantitativa)
distribuzione non omogenea

Produzione di proteine ricombinanti per studi strutturali
Spettrometria di massa: ionizzazione

SELDI (Surface Enhanced Laser Desorbtion Ionization)

👍 campioni biologici

Produzione di proteine ricombinanti per studi strutturali
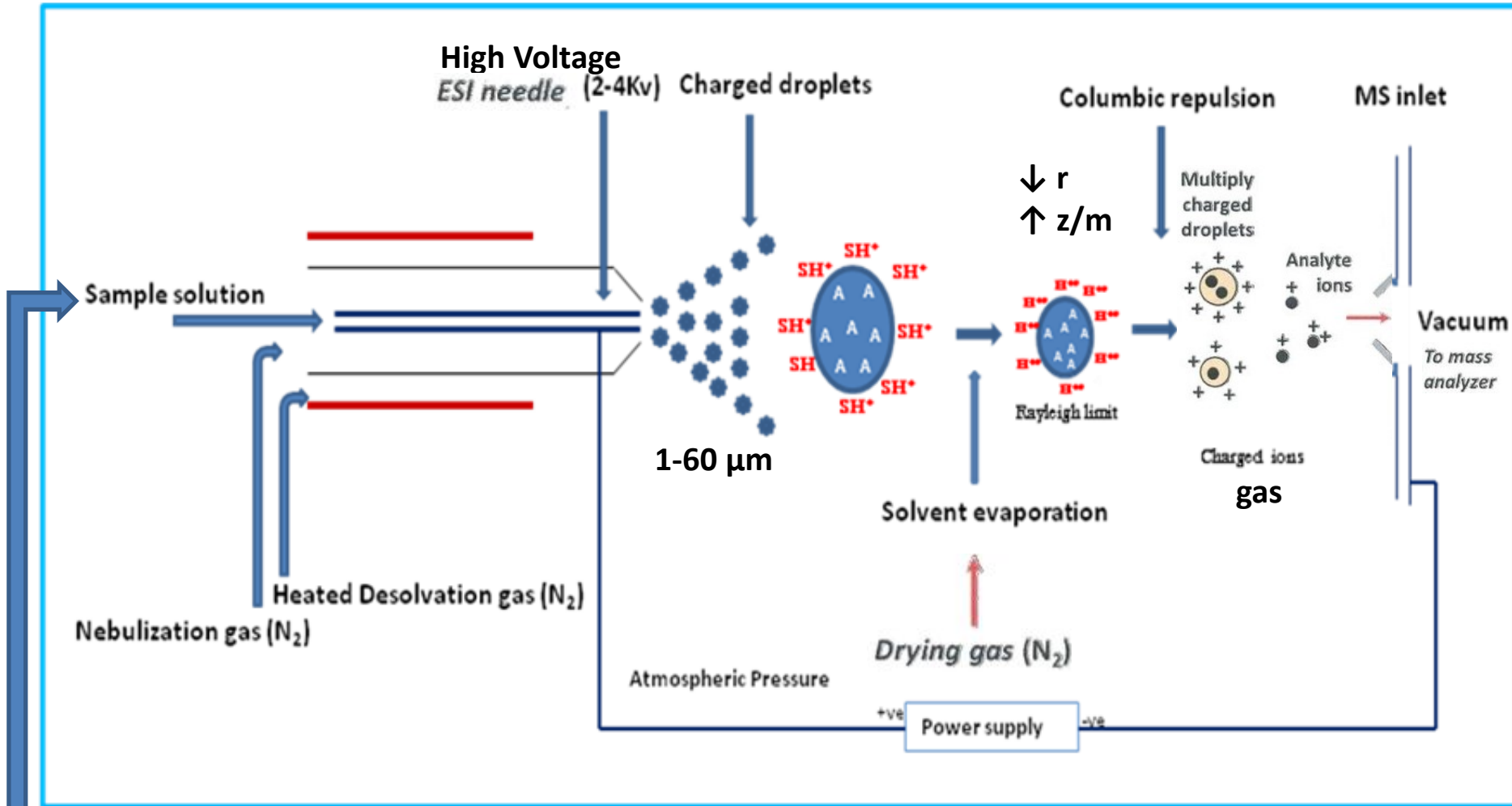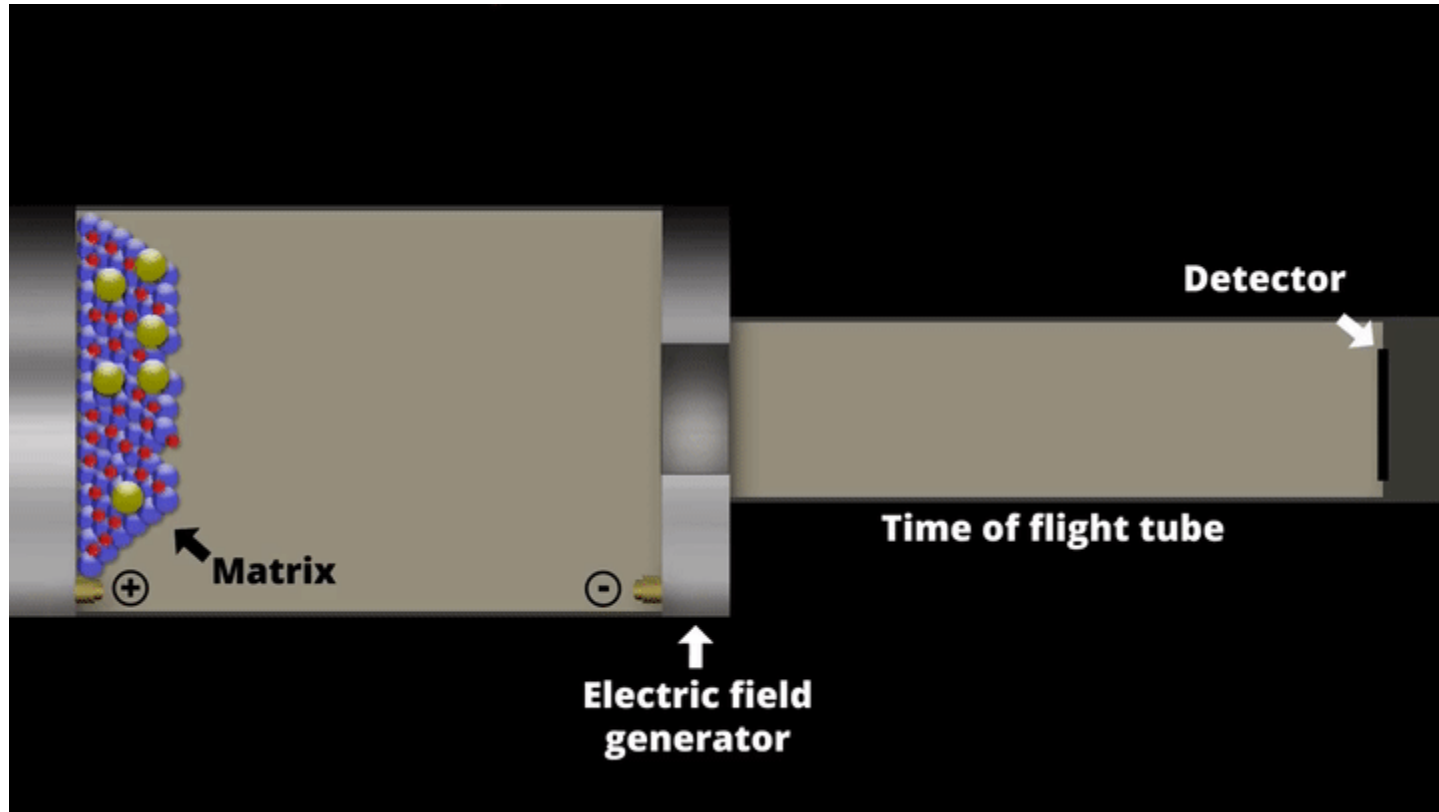## Spettrometria di massa: ionizzazione

ESI (ElectroSpray Ionization)
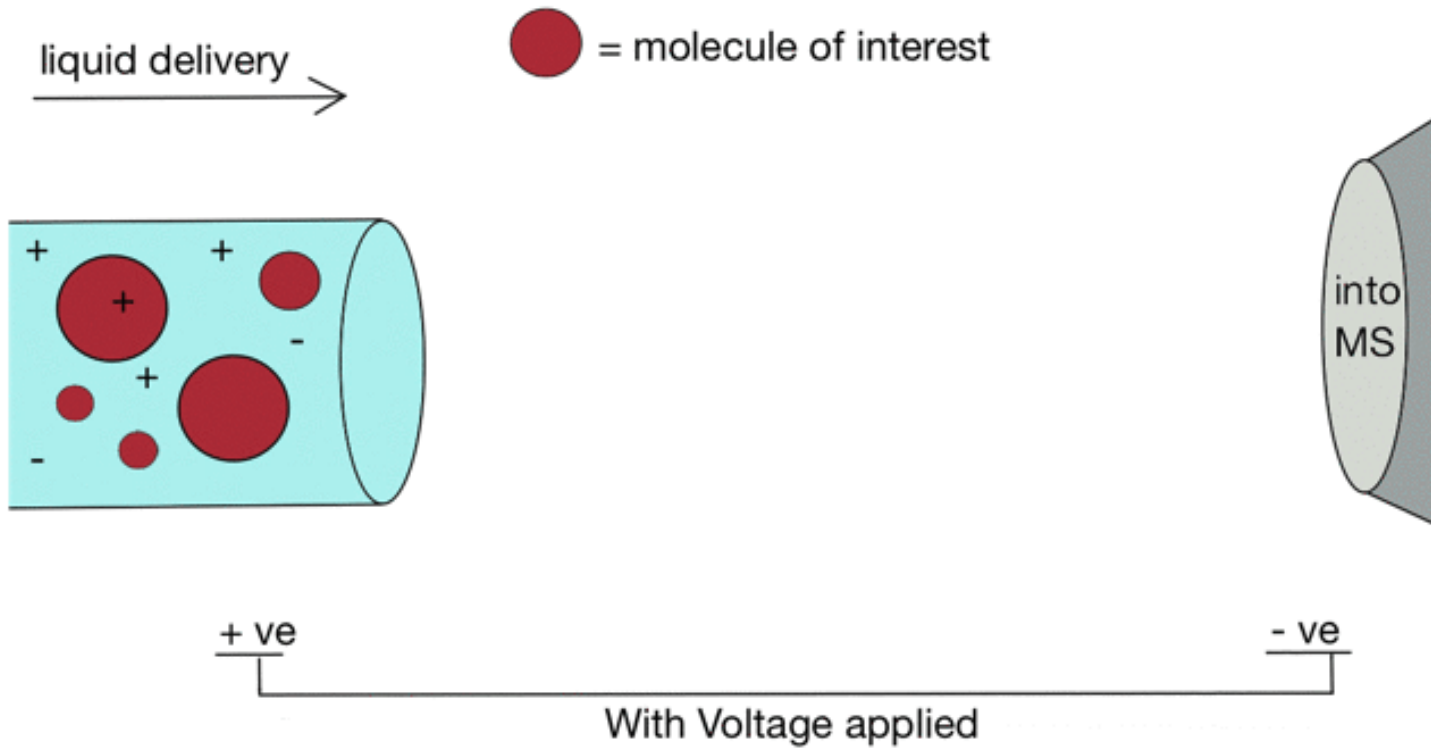
Produzione di proteine ricombinanti per studi strutturali
Spettrometria di massa: ionizzazione

ESI (ElectroSpray Ionization)

HPLC   capillare (5-10 µl/min; r=0.25-0.33 mm)
nano (0.07-0.15 µl/min; r=0.075 mm)→cala r gocce di 100-1000x → 100% ionizzazione

Produzione di proteine ricombinanti per studi strutturali
Spettrometria di massa: analizzatori

TOF (Time Od Fly)

Produzione di proteine ricombinanti per studi strutturali
Spettrometria di massa: analizzatori

TOF (Time Od Fly)

Source Region

Field Free Region (Analyzer)

Detector

Target   Accelaration Grid

electric field   drift region
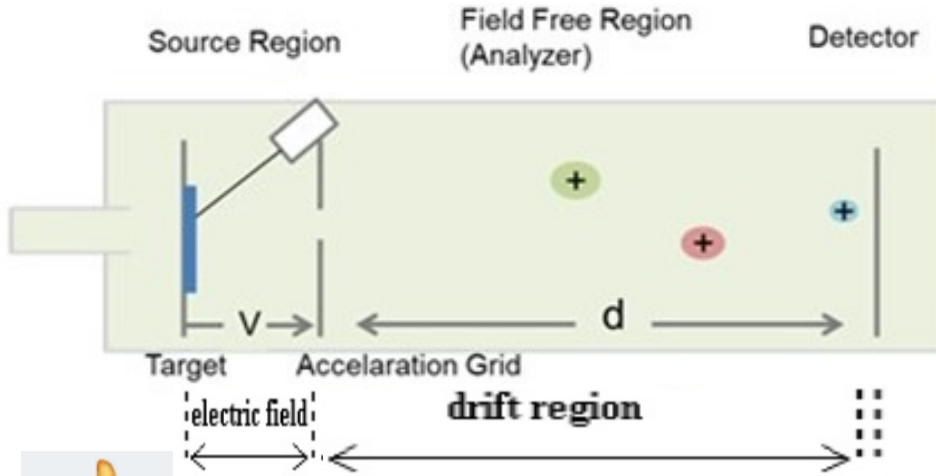
MW 60-80 kDa

Low resolution

The potential energy :
$$E=zeV$$
The kinetic energy of ions:
$$E=1/2mv^2$$
The potential energy is converted into kinetic energy:
$$zeV=1/2mv^2$$
Thus,

$$v = \sqrt{\frac{2zeV}{m}}$$

The time taken by ions to reach the detector: $t=\frac{d}{v}$

Finally, we get:

$$t= d\sqrt{\frac{m}{2zeV}}$$

Reflection

Detector1

Detector2
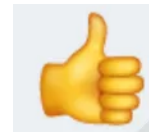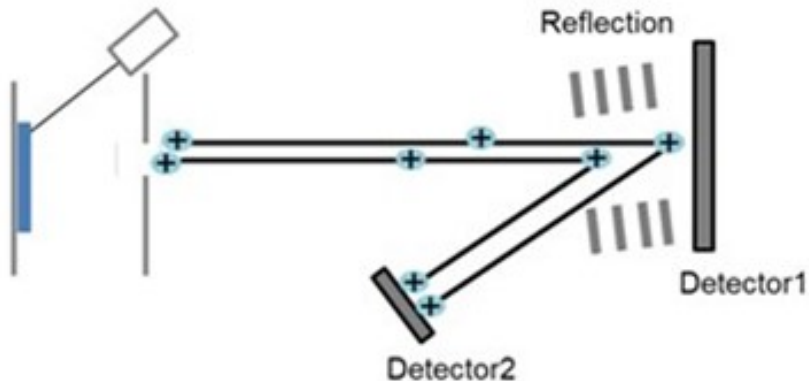
MW 5-10 kDa

High resolution

Produzione di proteine ricombinanti per studi strutturali
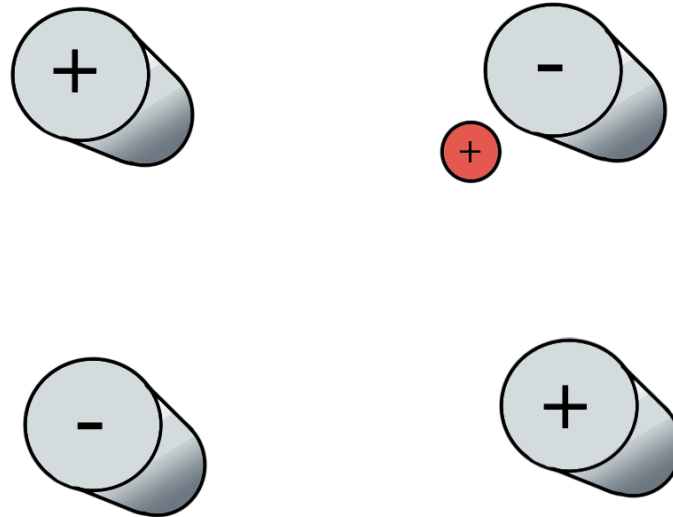## Spettrometria di massa: analizzatori

Quadrupole mass analyser
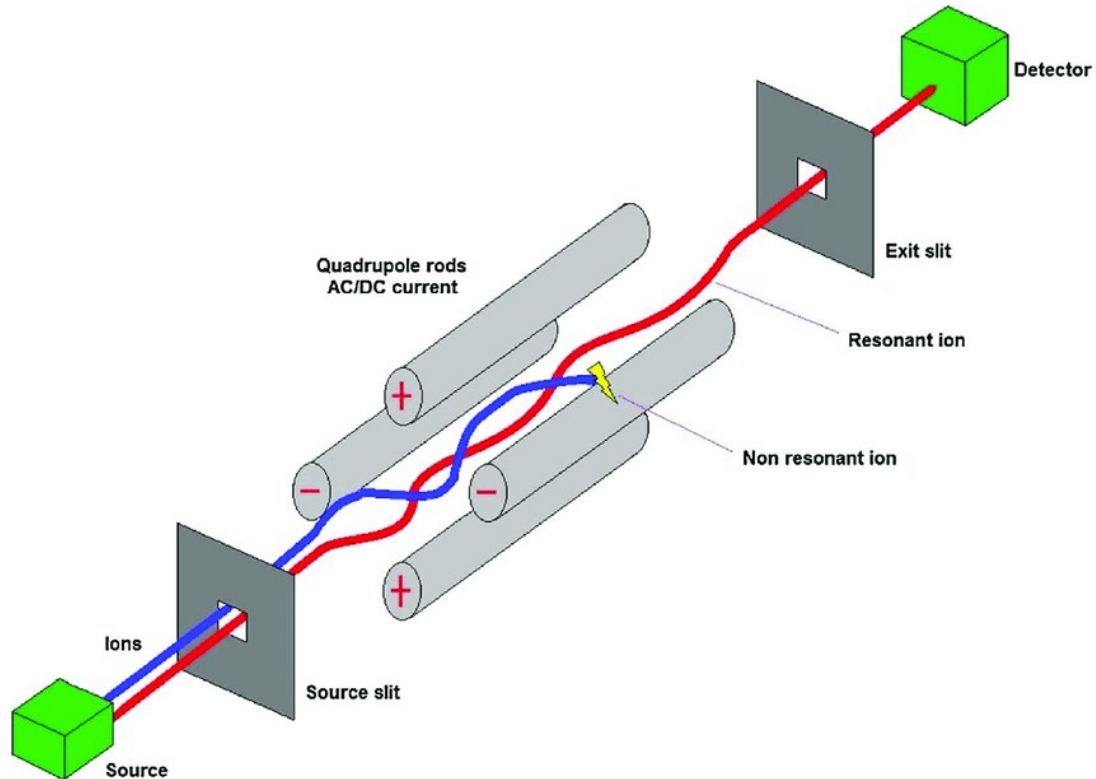
Produzione di proteine ricombinanti per studi strutturali
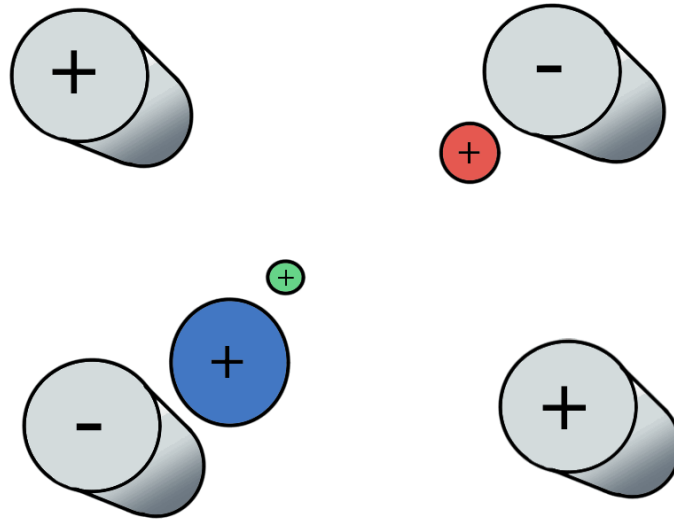## **Spettrometria di massa: analizzatori**

Quadrupole mass analyser

Per controllare come si muove uno ione nello spazio: Quattro poli.
Ogni ione con un dato m/z avrà un percorso stabile attraverso il quadripolo in date condizioni (determinato dalle ampiezze RF e DC).

Spettrometria di massa: analizzatori

Quadrupole mass analyser

Produzione di proteine ricombinanti per studi strutturali
Spettrometria di massa: analizzatori

Quadrupole mass analyser
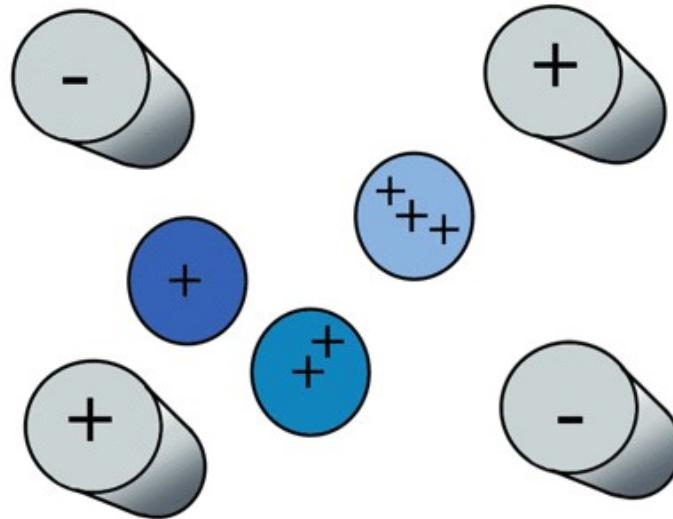
m / z determina il modo in cui uno ione si muove in un campo elettromagnetico.
Stesso ione blu presente in tre diversi stati di carica:
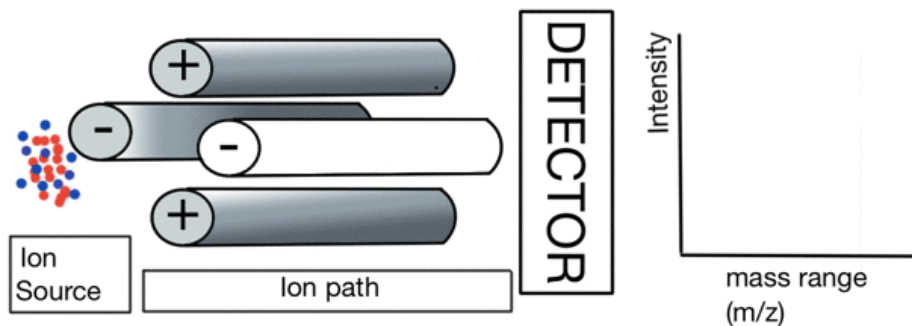2+ si comporterebbero come se avessero la metà della massa
3+ sembrerebbero avere un terzo della massa dello ione 1+.

Produzione di proteine ricombinanti per studi strutturali
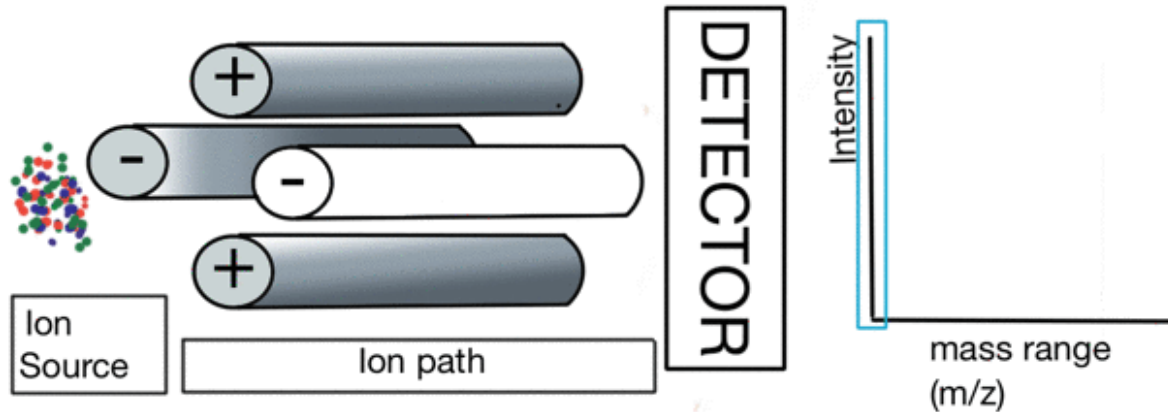## Spettrometria di massa: analizzatori

Quadrupole mass analyser
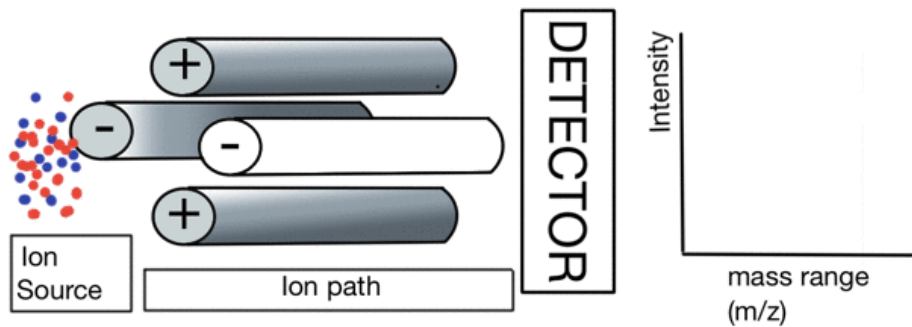
Produzione di proteine ricombinanti per studi strutturali
Spettrometria di massa: analizzatori

Quadrupole mass analyser

The quadrupole SCANS through a range of conditions...

… that lets ions of a specific mass to charge ratio reach the detector.

All the other ions go to waste (over a few milliseconds).

spettro cambiando sequenzialmente dell'ampiezza RF del quadrupolo per consentire a diversi ioni di raggiungere il rivelatore.

👍 Spettri di massa classici
Riproducibilità
Sistemi relativamente piccoli ed economici.

👎 Risoluzione limitata
Altezze dei picchi variabili in funzione di m

Produzione di proteine ricombinanti per studi strutturali
Spettrometria di massa: analizzatori

Single Focusing Magnetic Mass Analyzer
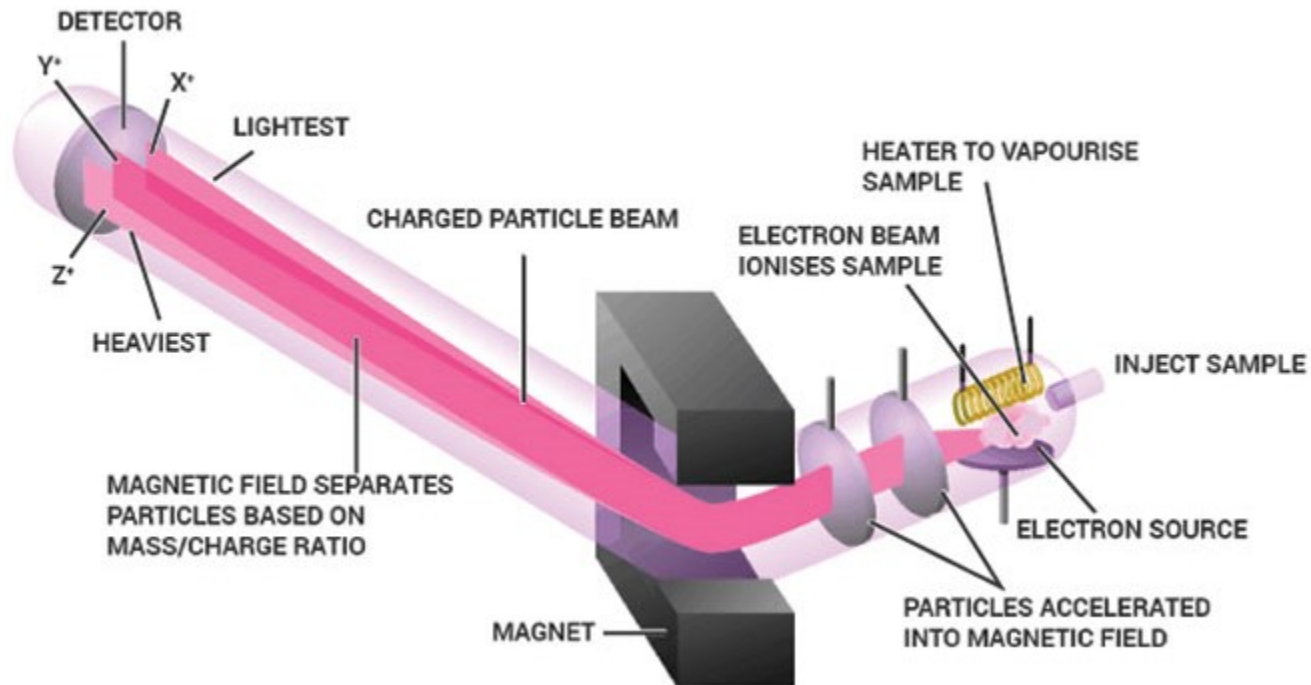
tubo di vetro a forma di ferro di cavallo sotto vuoto

ingresso campione,
sorgente di bombardamento di elettroni,
 piastre di accelerazione su un'estremità
collettore all'altra estremità.

Alla curvatura del tubo l'applicazione di un campo elettrico / magnetico.
Il campione sotto forma di vapore è consentito attraverso l'ingresso e
bombardato con un fascio di elettroni a 70eV.

ioni altamente carichi +

accelerate dalle placche

un percorso rettilineo.

Applicando un campo elettrico o magnetico,
viaggiano lungo un percorso curvo
gli ioni separati in base alle loro masse

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Abbassamento della barriera entropica
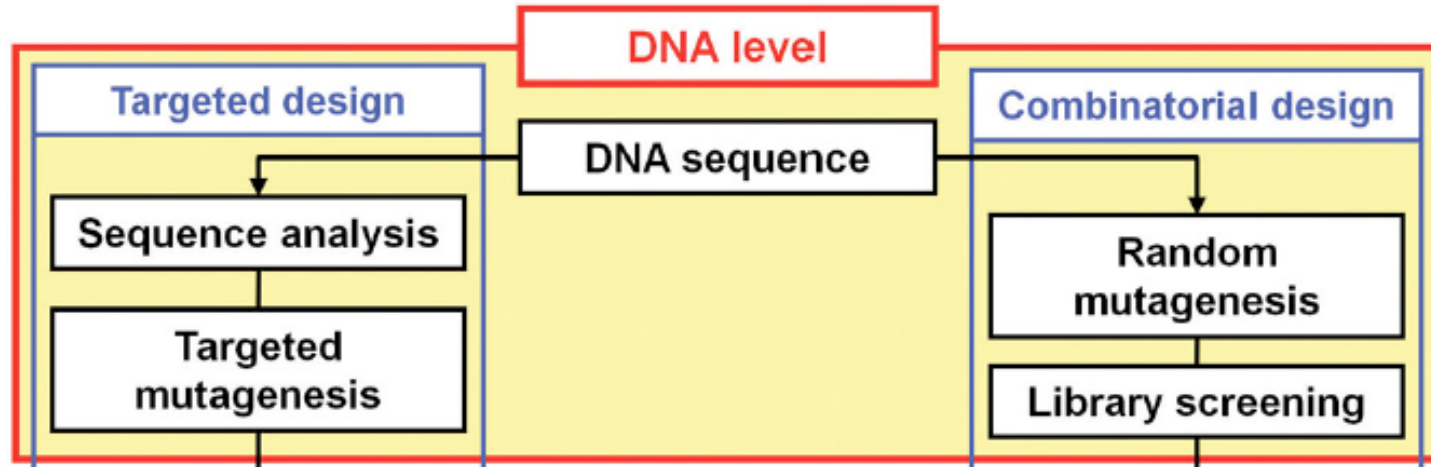
1. **Mutanti di delezione**
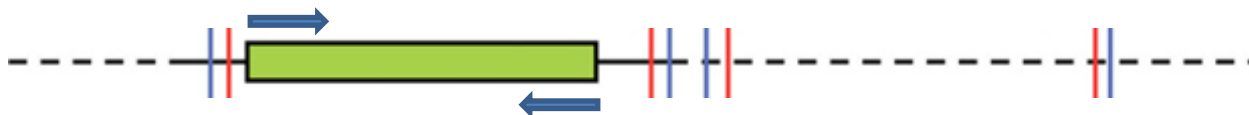
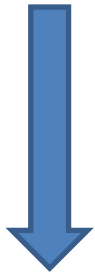**2.**   **Generazione di libraries di mutanti di delezione mediante:**
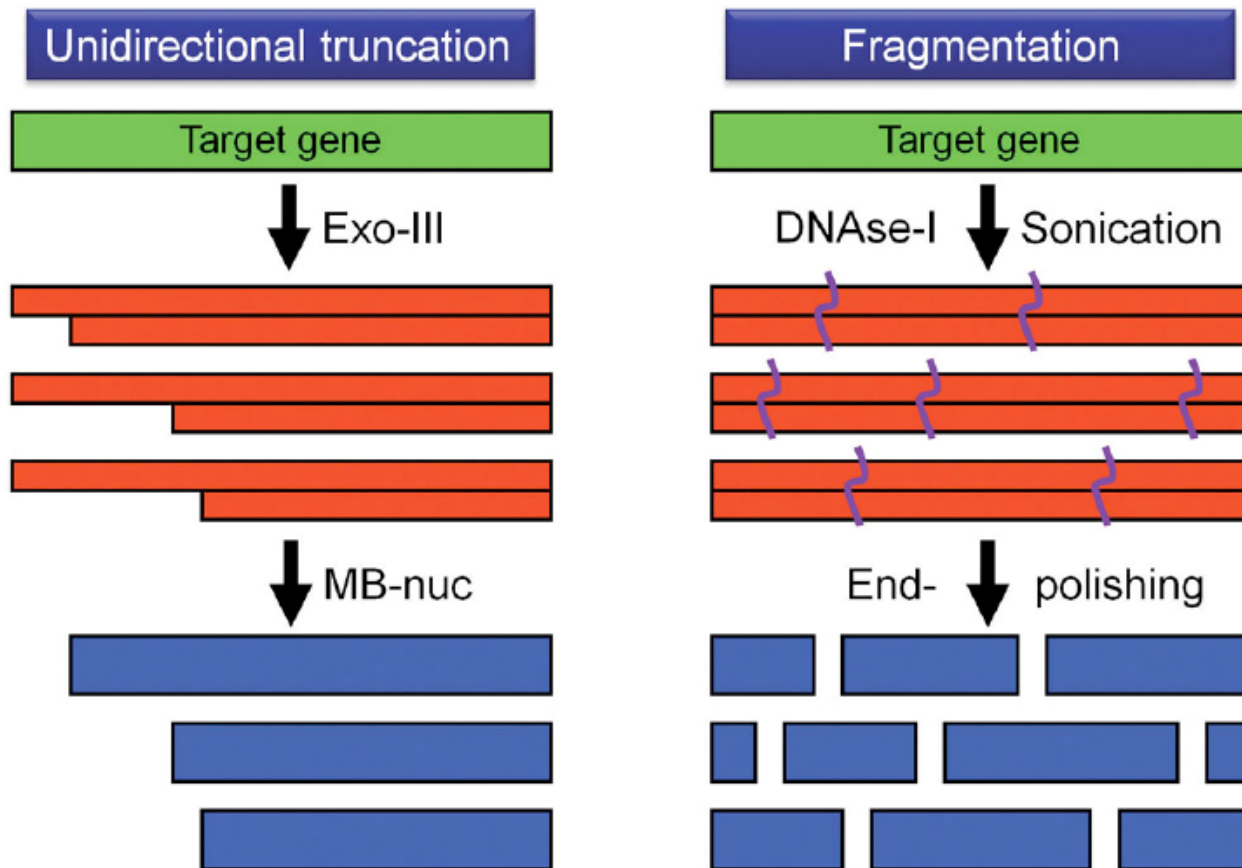
Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Abbassamento della barriera entropica

- **Surface Entropy Reduction:**   **MODIFICAZIONI CHIMICHE:**

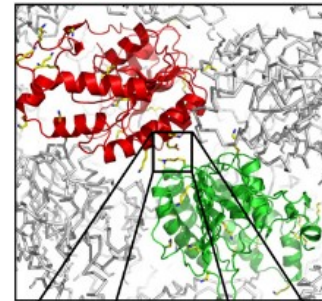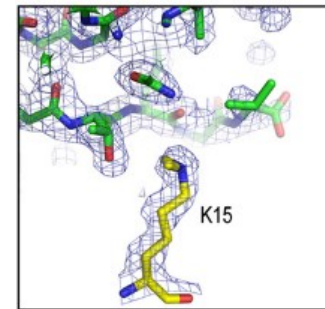**1)   LYSIN REDUCTIVE METHYLATION**

K superficiali

metilate

interazioni idrofobiche

Lysine Methylation as a Routine   Ways & Means
Rescue Strategy for Protein Crystallization

Thomas S. Walter,[1,3] Christoph Meier,[1,2,3,4]
Rene Assenberg,[1,3] Kin-Fai Au,[1,2,3,4]
Jingshan Ren,[1,2,3,4] Anil Verma,[1,3]
Joanne E. Nettleship,[1,3] Raymond J. Owens,[1,3]
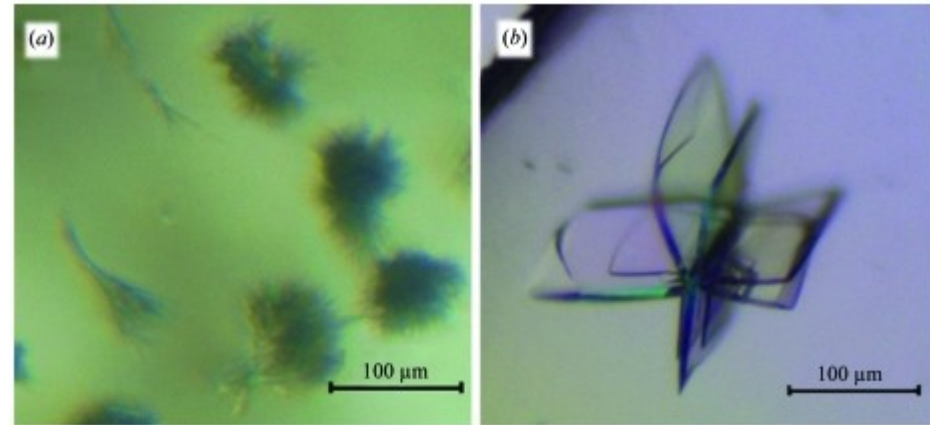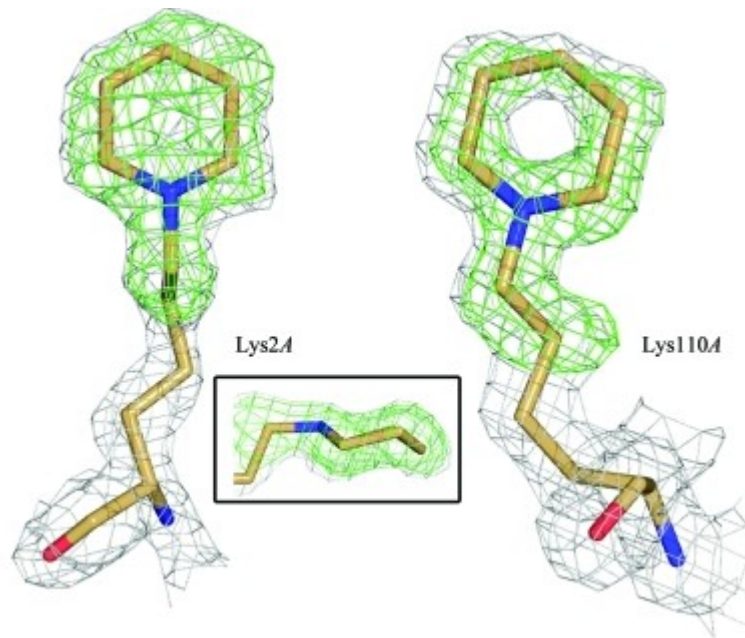David I. Stuart,[1,2,3,4] and Jonathan M. Grimes[1,2,3,4,*]

K15

E166

K177
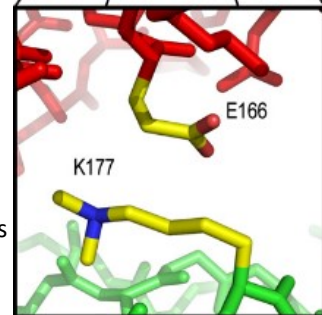
Examples of Methylated Lysines
from MVE Methyltransferase

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Abbassamento della barriera entropica

- **Surface Entropy Reduction: MODIFICAZIONI CHIMICHE:**

2) **REDUCTIVE CYCLIC PENTYLATION OF FREE AMINES**

Crystallization of a pentapeptide-repeat protein by reductive cyclic pentylation of free amines with glutaraldehyde
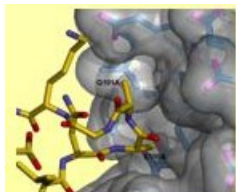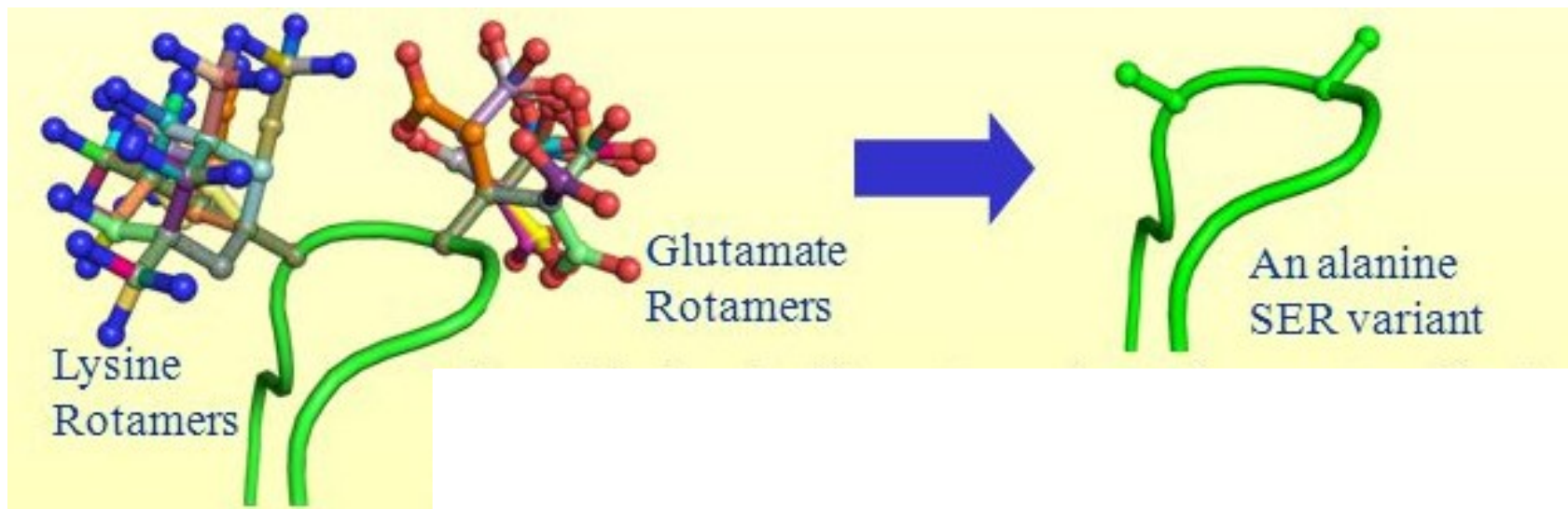
Matthew W. Vetting,[a,*] Subray S. Hegde,[a] and John S. Blanchard[a]

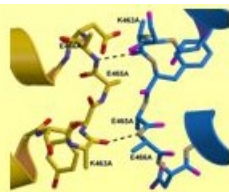# Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Abbassamento della barriera entropica

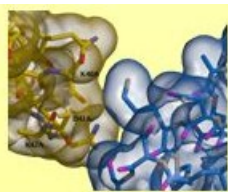• **Surface Entropy Reduction: biologia molecolare**
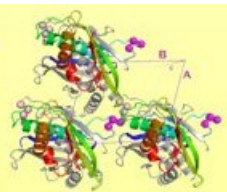
MUTAGENESI sito specifica

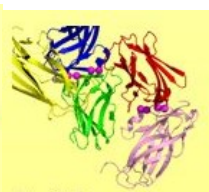Lysine Rotamers — Glutamate Rotamers → An alanine SER variant

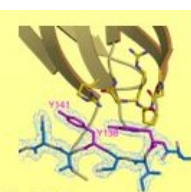

Hsp33
Structure 12:1901

RGSL of PDZ-RhoGEF
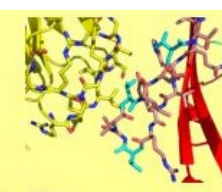Structure 9:559

LcrV
Structure 12:357

OhrB
Acta Cryst D63:1269

RhoGDI
K99S, Q100S
Acta Cryst D63:636

RhoGDI
K138Y, K141Y
Acta Cryst D63:636

RhoGDI
E155H, E157H
Acta Cryst D63:636

RhoGDI
K135T, K138T, K141T
Acta Cryst D63:636

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Abbassamento della barriera entropica

MUTAGENESI sito specifica: selezione dei residui I

http://services.mbi.ucla.edu/SER/

PROTEIN|SCIENCE

research-article | Free Access

Toward rational protein crystallization: A Web server for the design of crystallizable protein variants

Lukasz Goldschmidt, David R. Cooper, Zygmunt S. Derewenda, David Eisenberg

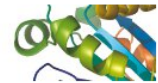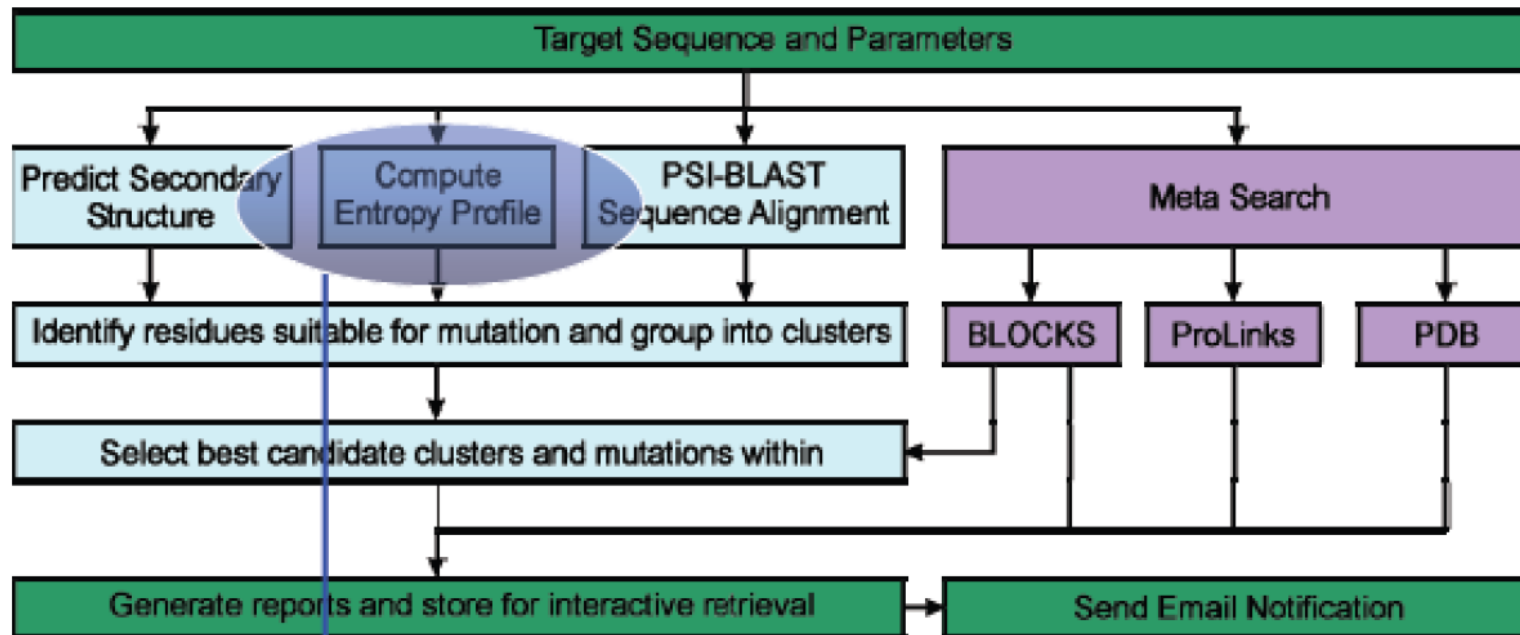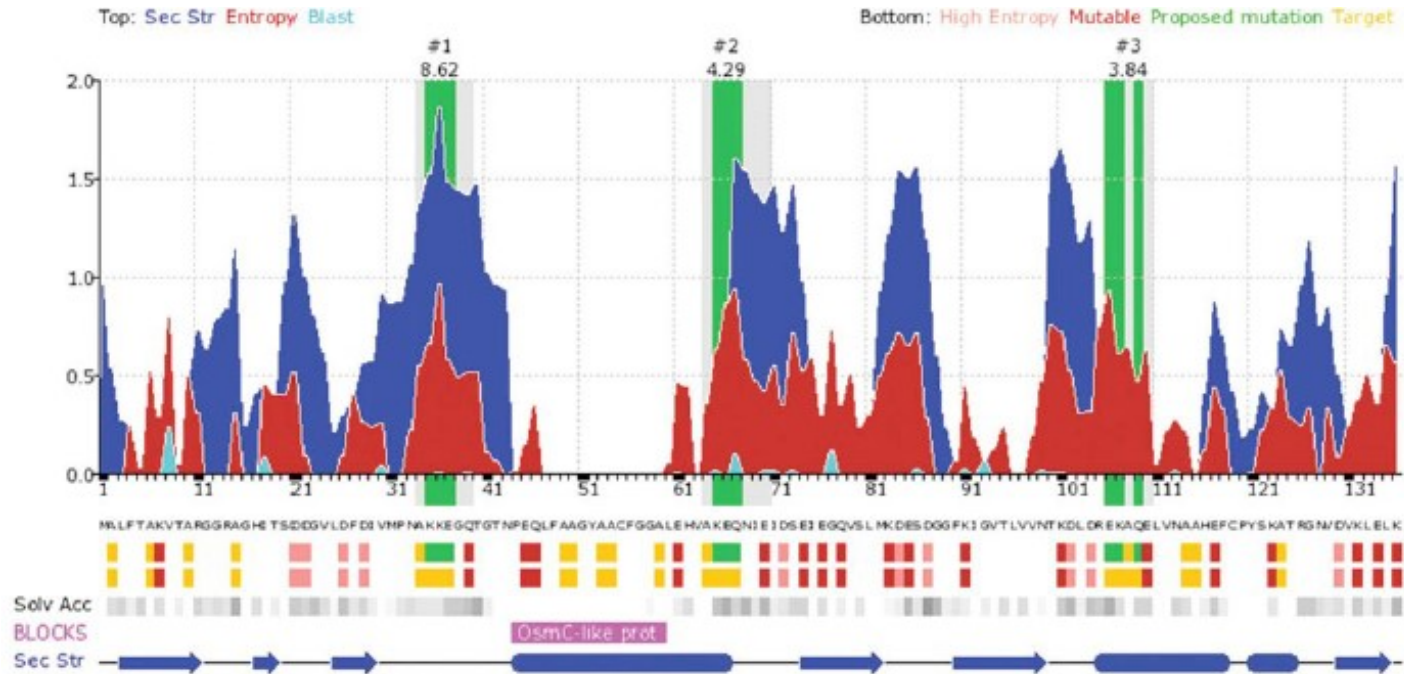First published: 02 January 2009 | https://doi.org/10.1110/ps.072914007 | Cited by: 134

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Abbassamento della barriera entropica

MUTAGENESI sito specifica: selezione dei residui II
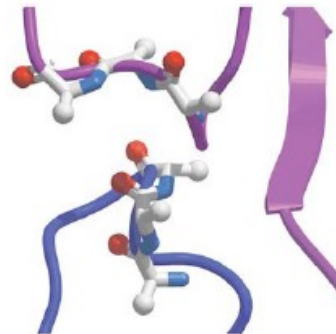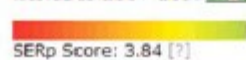
Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: Abbassamento della barriera entropica

• **Surface Entropy Reduction: MUTAGENESI sito specifica:**

non solo A, ma anche Y, H e S favoriscono interazioni intermolecolari
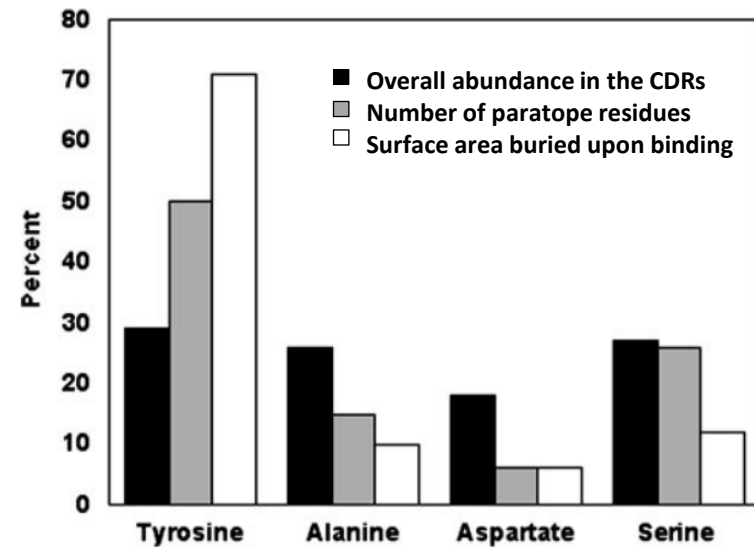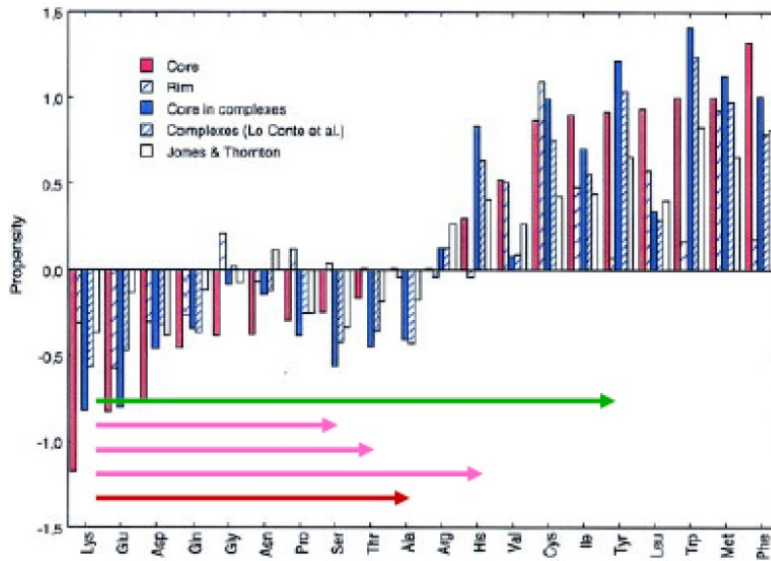
Proteins: Structure, Function, and Bioinformatics / Volume 53, Issue 3

Dissecting subunit interfaces in homodimeric proteins

Ranjit Prasad Bahadur, Pinak Chakrabarti, Francis Rodier, Joël Janin ✉

Synthetic antibodies from a four-amino-acid code:
A dominant role for tyrosine in antigen recognition

Frederic A. Fellouse*, Christian Wiesmann*, and Sachdev S. Sidhu*†

Produzione di proteine ricombinanti per studi strutturali
Strategie di improvement: summary
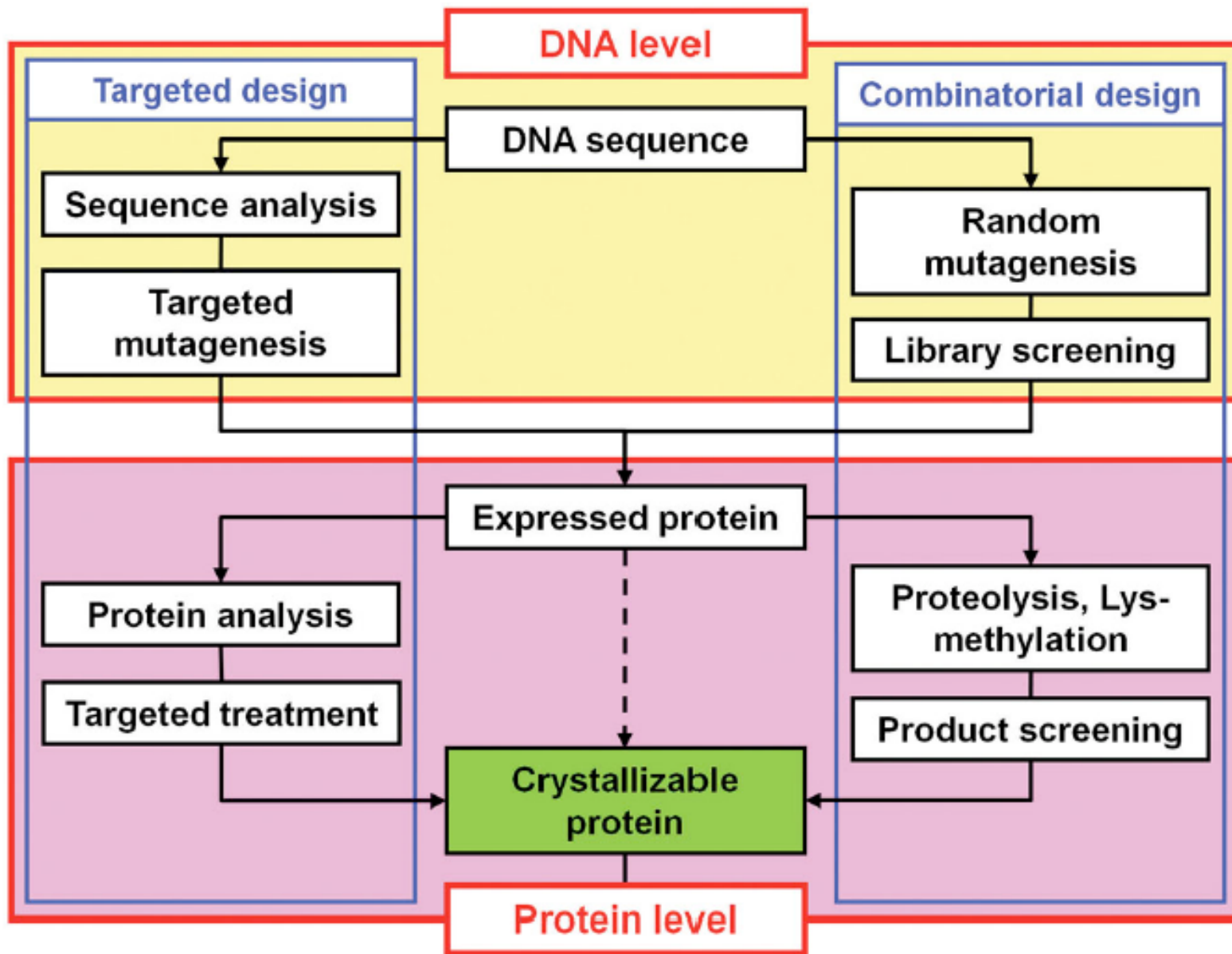
Produzione di proteine ricombinanti per studi strutturali
Cristallizzazione: probabilità di successo
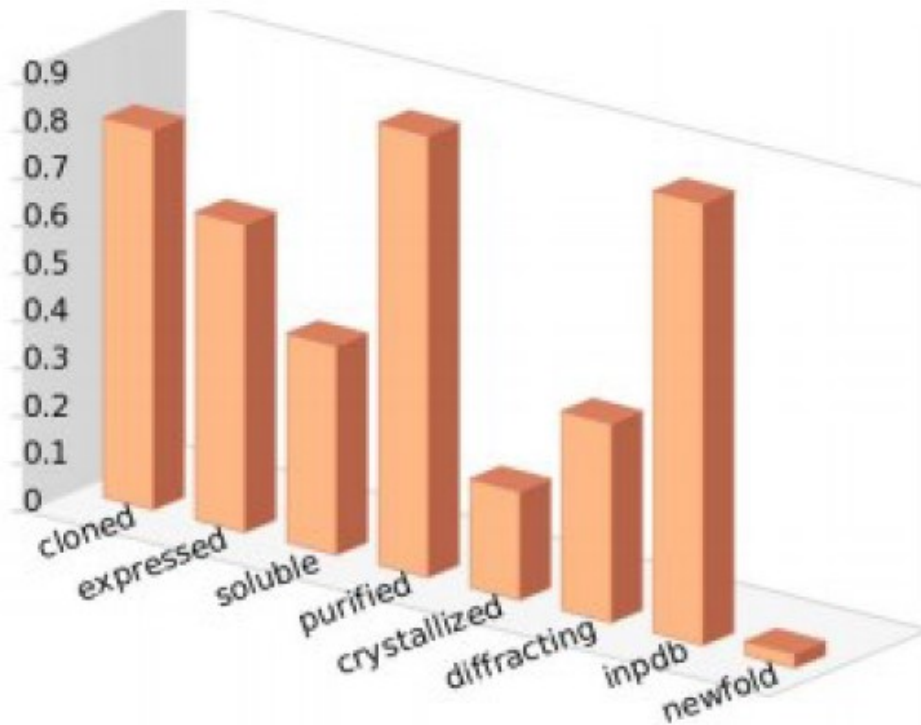
Classi definite per diversa percentuale di successo nella cristallizzazione osservata in TargetDB (database centralizzato dai centri di genomica strutturale)

Espressione di proteine ricombinanti per studi strutturali
Predizione del successo della cristallizzazione

http://xtalpred.godziklab.org/XtalPred-cgi/xtal.pl
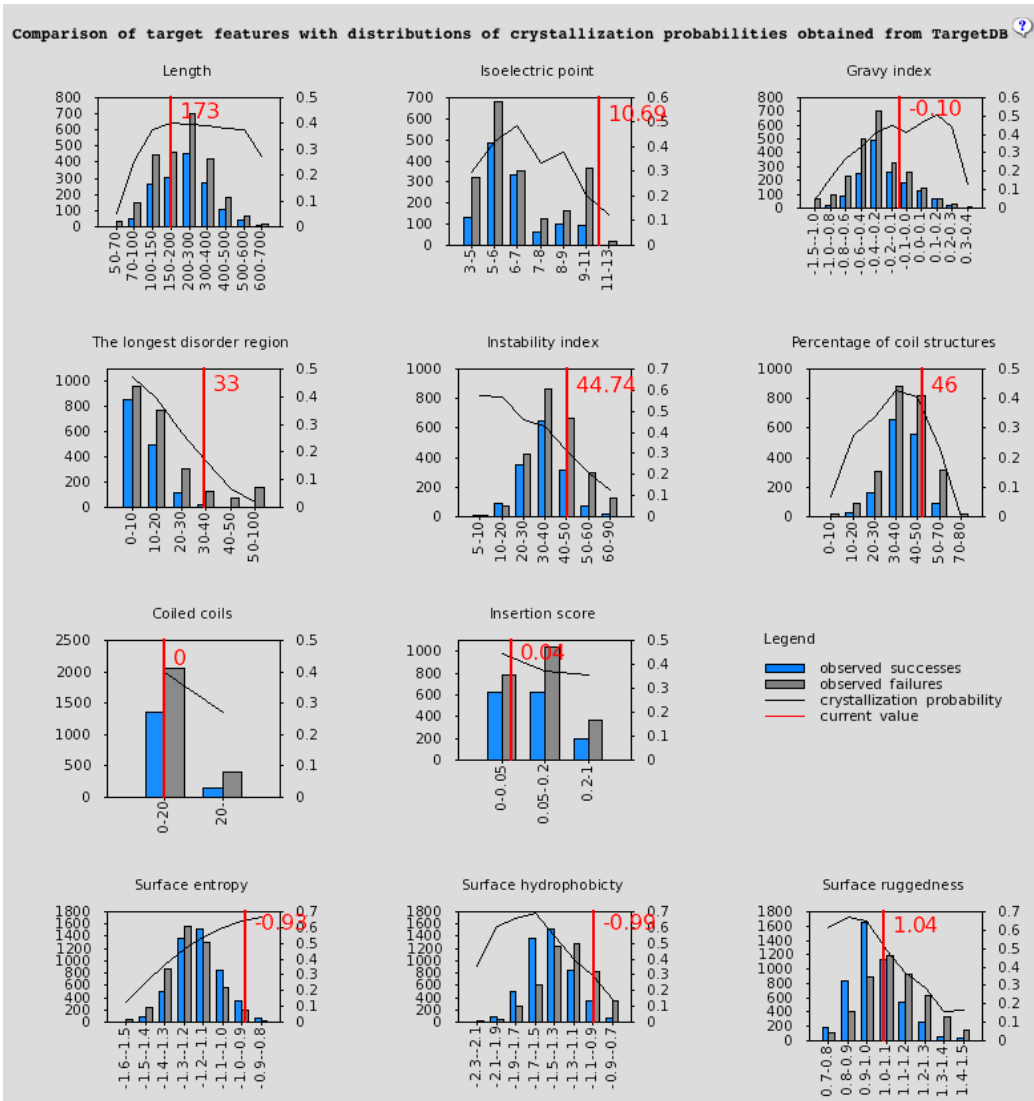
Combinando le probabilità individuali di cristallizzazione calcolate per le classi di parametri,

il target viene assegnato a una delle cinque classi di cristallizzazione.

Espressione di proteine ricombinanti per studi strutturali
Predizione del successo della cristallizzazione

Crystallization class: 2

Crystallization classes

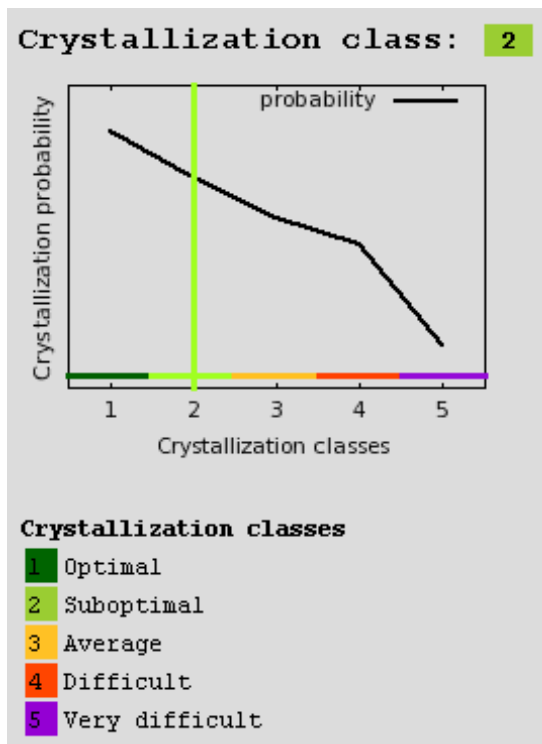|	
1	Optimal
2	Suboptimal
3	Average
4	Difficult
5	Very difficult

RESULTS
- observed successes
- observed failures

TARGET CLASS
- optimal
- suboptimal
- average
- difficult
- very difficult

#	Target id (link to target details)	EP-class	RF-class	Length	Gravy index	Instability index (II)	Isoelectric point (pI)	Coiled coils	Longest disorder region	Percentage of coil structure	Transmembrane helices (TM)	Signal peptides (SP)	Insertions score	Homologs in NR (clustered to 60% seq. ident.)	Homologs in PDB
1	sp|Q14118	5	11	624	-0.39	45.65	9.31	0	234	69	No	No	0.05	871	227

EP crystallization classes
↓ Most promising
1 2 3 4 5
↑ Least promising

RF crystallization classes
↓ Most promising
1 2 3 4 5 6 7 8 9 10 11
Least promising ↑