Ingénierie des protéines

Stéphane Delbecq
EA 4558 Vaccination antiparasitaire
Laboratoire de Biologie Cellulaire et Moléculaire
Faculté de Pharmacie - Montpellier

Rappel: transcription et traduction

Universalité du code génétique et de la molécule d'ADN

Biologie moléculaire et génie génétique

Production de protéines recombinantes

But de l'ingénierie des protéines:

- produire une protéine à volonté (temps, quantité, prix…)

- produire une protéine fonctionnelle

- modifier les propriétés d'une protéine (stabilité, immunogénicité, activité enzymatique)

Méthodes classiques: clonage, mutagénèse

Evolution dirigée des protéines

Méthodes classiques

Clonage moléculaire : insertion d'une séquence d'ADN dans un vecteur (plasmide, phage…) puis introduction dans une cellule-hôte (bactérie, mammifère…)

Caractéristiques d'une séquence codante utilisable en expression:

- pas d'intron
- pas de séquences signales (sauf nécessité)
- en phase avec les tags ou protéines de fusion

Différentes méthodes:

- restriction/ligation
- PCR
- recombinaison *in vivo* ou *in vitro*

Choix de la séquence à cloner

Le départ:

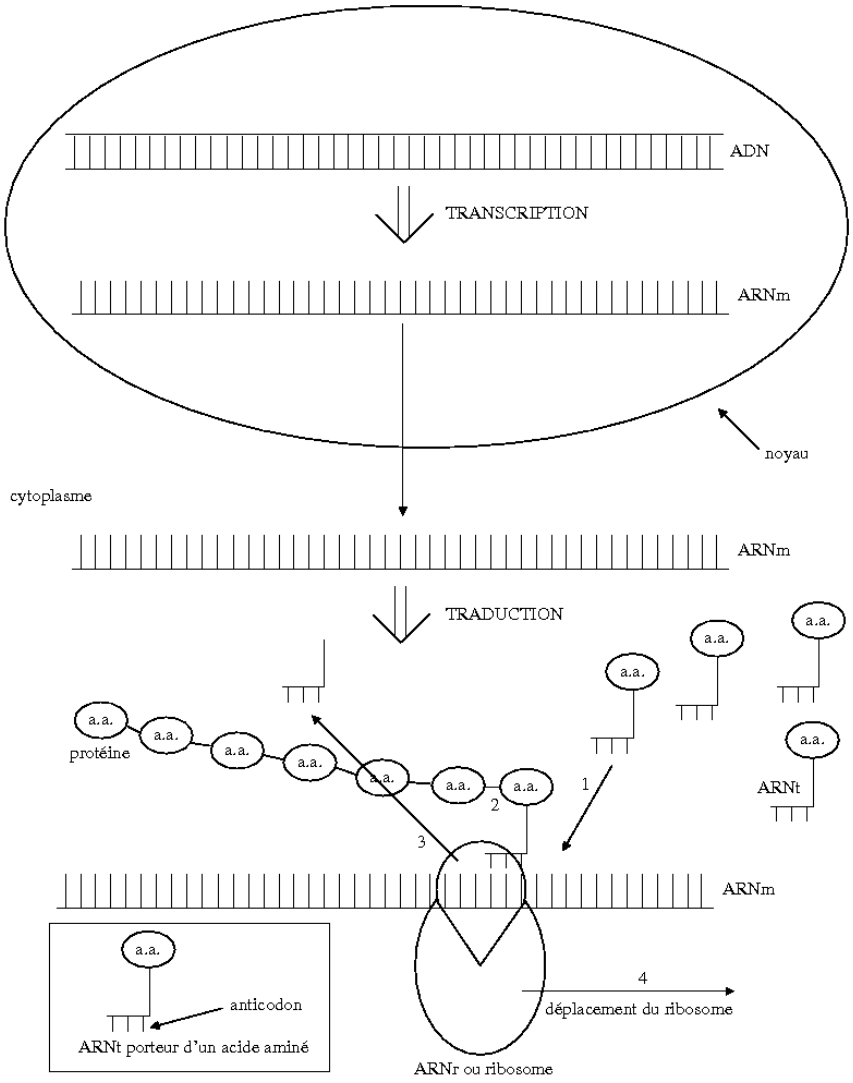
- séquence déjà clonée dans un vecteur (ex: clone isolé d'une banque d'ADN)
  = sous-clonage
- séquence connue mais pas de clone disponible
  = clonage à partir de l'organisme d'origine
- séquence inconnue mais existence d'homologue dans d'autres organismes (ex: clonage d'une enzyme de parasite présente également chez l'homme)
  = recherche de la séquence (PCR dégénérée ou séquençage du génome)
- séquence inconnue sans homologue
  = recherche du gène à partir de la protéine (génération d'anticorps puis criblage de banques ou spectrométrie de masse)

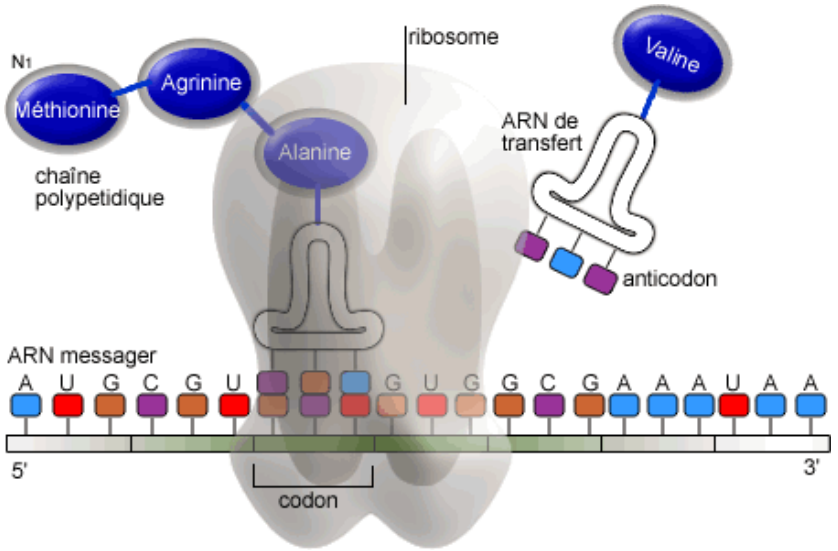
Origine de la séquence:

- organisme d'origine : procaryote ou eucaryote
- ADN d'origine: vecteur, génomique ou ADNc?

Définir la stratégie de clonage à adopter !

La synthèse des protéines chez les procaryotes



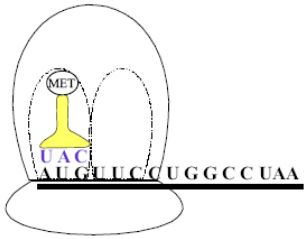


Traduction co-traductionelle

ARNm : RBS, AUG et STOP

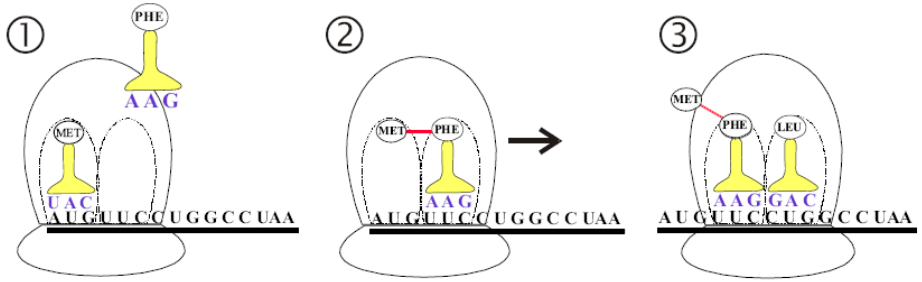
La traduction a lieu en 3 phases :

La phase d'initiation



- Fixation du ribosome au niveau du codon initiateur AUG. Il signale le point dedépart de la traduction.
- Début de la lecture de l'ARNm, le codon AUG code pour la méthionine (Met). Fixation de l'ARNt-Met.

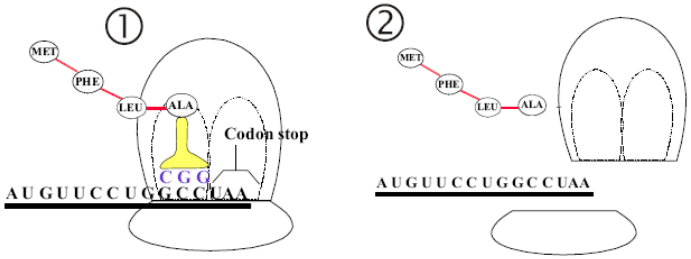
La phase d'élongation



Lecture successive des codons et formation du polypeptide :
→ ① lecture du codon situé dans le site A du ribosome
→ ① fixation de l'ARNt-acide aminé correspondant
→ ② formation d'une liaison peptidique avec le polypeptide présent dans le site P
→ ③ le ribosome se déplace d'un codon.
→ ① lecture du codon situé dans le site A du ribosome...

La phase de lecture ouverte est définie par le codon AUG

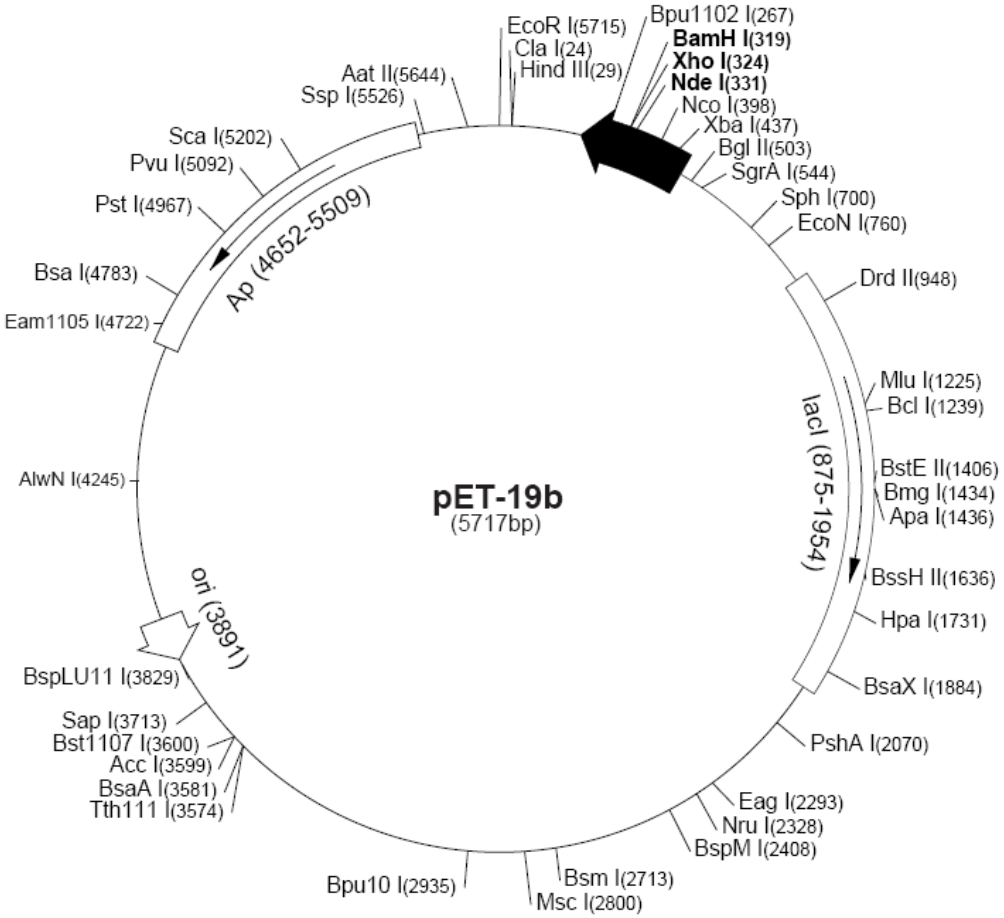
La phase de terminaison



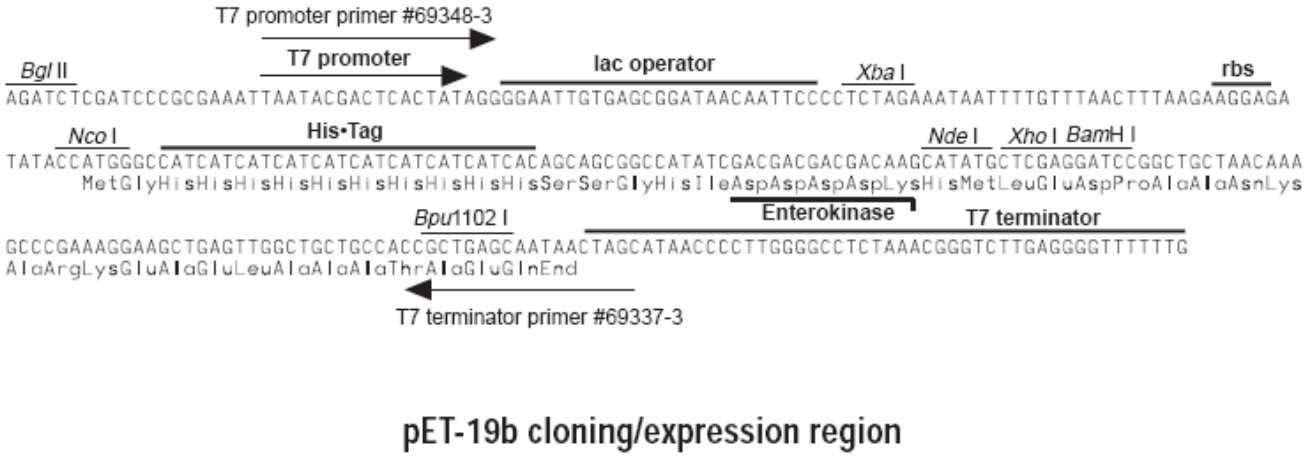
①Le ribosome arrive à un codon stop ou "non-sens" qui ne correspond à aucun ARNt. La traduction est bloquée.
②Le ribosome se détache, le polypeptide est libéré.

Caractéristique d'un vecteur d'expression (*E. coli*)

pET-19b sequence landmarks	
T7 promoter	472-488
T7 transcription start	471
His• Tag coding sequence	366-395
Multiple cloning sites	
(*Nde* I -*Bam*H I)	319-335
T7 terminator	213-259
lacI coding sequence	875-1954
pBR322 origin	3891
bla coding sequence	4652-5509



Caractéristique d'un vecteur d'expression (*E. coli*)



Promoteur: séquence de fixation de l'ARN polymérase (ici T7 pol)
Rbs: ribosome binding site
ATG: codon d'initiation (Met)
Enterokinase: site de clivage spécifique
T7 terminator: terminateur de transcription

Cassette d'expression : contient les éléments permettant la synthèse d'un ARNm codant pour la protéine d'interêt

Exemple d'une fusion entre 2 séquences codantes

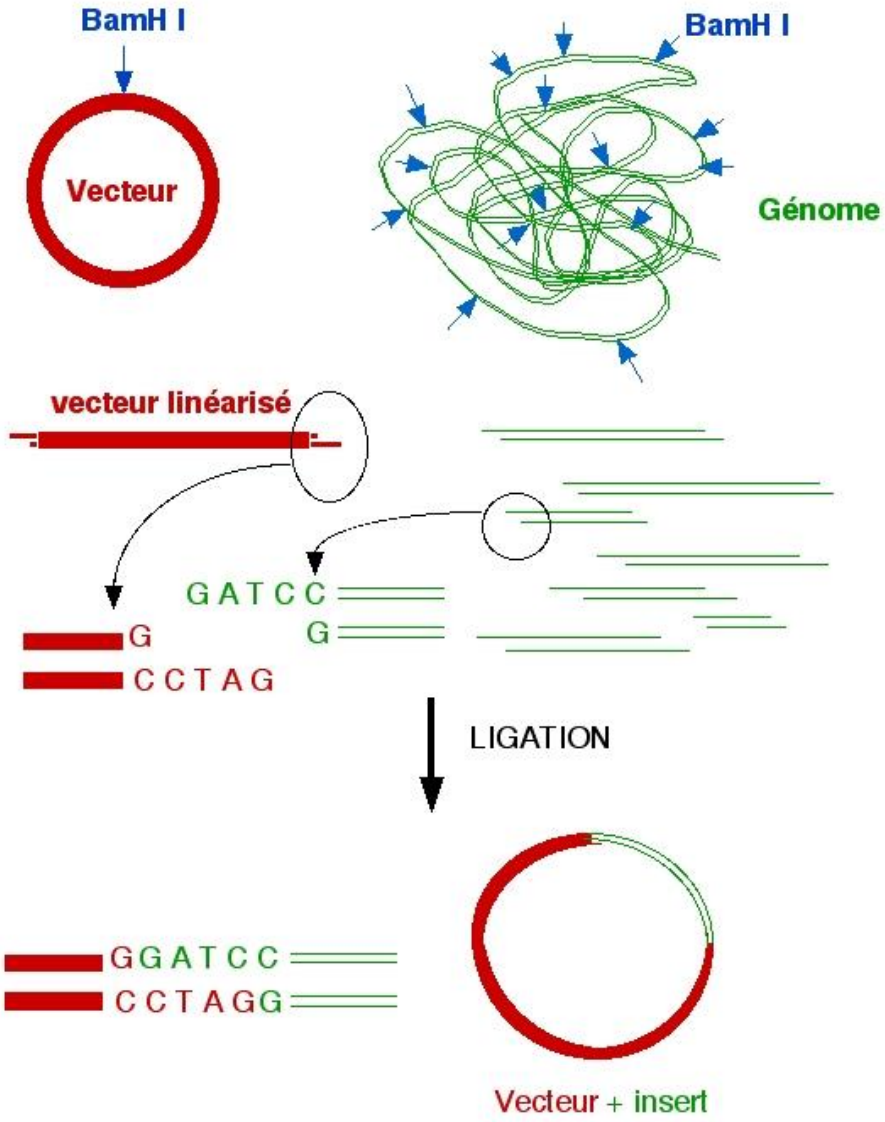
```
 M   A   S   S   L   R   Q   I   L   D   S   Q   K   M   E   W   R   S      18
ATG GCT TCT TCT CTG CGT CAA ATT CTG GAT TCT CAA AAA ATG GAA TGG CGT TCT     54

 N   A   G   G   S   M   S   D   K   I   I   H   L   T   D   D   S   F      36
AAC GCT GGT GGT TCT ATG AGC GAT AAA ATT ATT CAC CTG ACT GAC GAC AGT TTT    108

 D   T   D   V   L   K   A   D   G   A   I   L   V   D   F   W   A   E      54
GAC ACG GAT GTA CTC AAA GCG GAC GGG GCG ATC CTC GTC GAT TTC TGG GCA GAG    162

 W   C   G   P   C   K   M   I   A   P   I   L   D   E   I   A   D   E      72
TGG TGC GGT CCG TGC AAA ATG ATC GCC CCG ATT CTG GAT GAA ATC GCT GAC GAA    216

 Y   Q   G   K   L   T   V   A   K   L   N   I   D   Q   N   P   G   T      90
TAT CAG GGC AAA CTG ACC GTT GCA AAA CTG AAC ATC GAT CAA AAC CCT GGC ACT    270

 A   P   K   Y   G   I   R   G   I   P   T   L   L   L   F   K   N   G     108
GCG CCG AAA TAT GGC ATC CGT GGT ATC CCG ACT CTG CTG CTG TTC AAA AAC GGT    324

 E   V   A   A   T   K   V   G   A   L   S   K   G   Q   L   K   E   F     126
GAA GTG GCG GCA ACC AAA GTG GGT GCA CTG TCT AAA GGT CAG TTG AAA GAG TTC    378

 L   D   A   N   L   A   *                                                 132
CTC GAC GCT AAC CTG GCC TAG
```

Rouge: peptide substrat de la biotine ligase BirA de *E. coli*
Noir: thioredoxine

→ **Expression d'une thioredoxine biotinylée *in vivo***

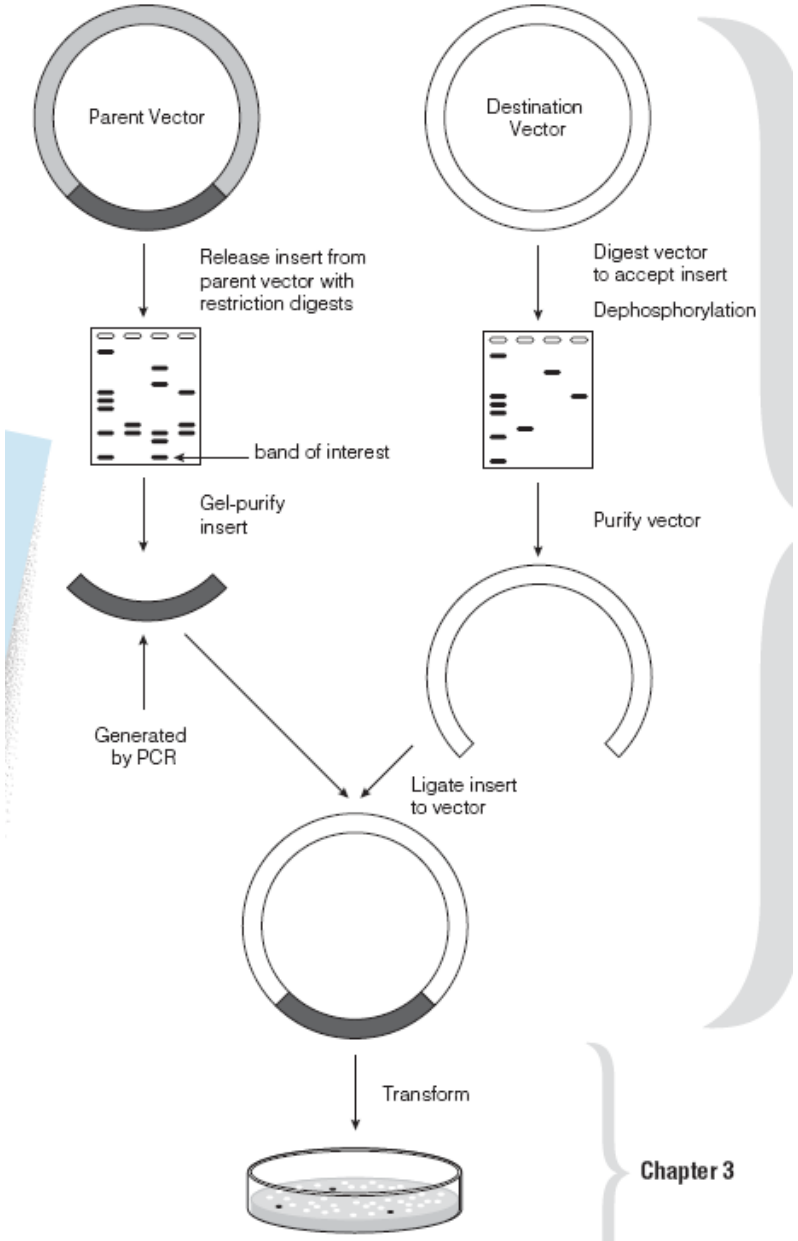
Méthode de clonage: restriction/ligation

Principe du clonage (cas d'une banque génomique)



Source : autre vecteur ou produit PCR

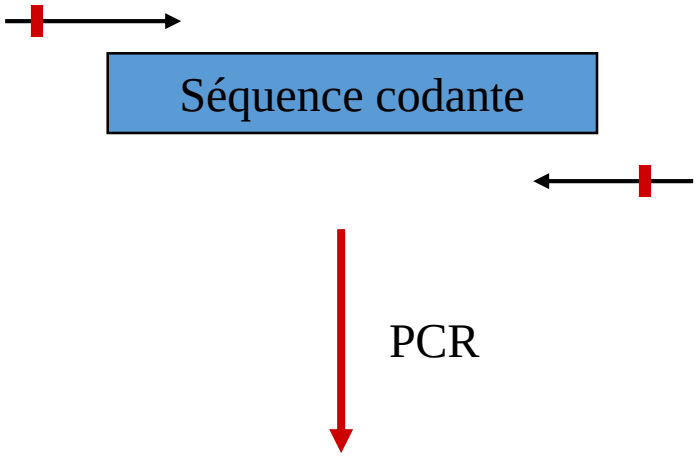
Méthode de clonage: restriction/ligation



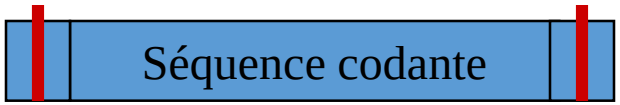
Sous-clonage direct : si les sites sont compatibles avec la phase de lecture ouverte

PCR: création des sites appropriés

Méthode de clonage: PCR



Site de restriction dans les amorces



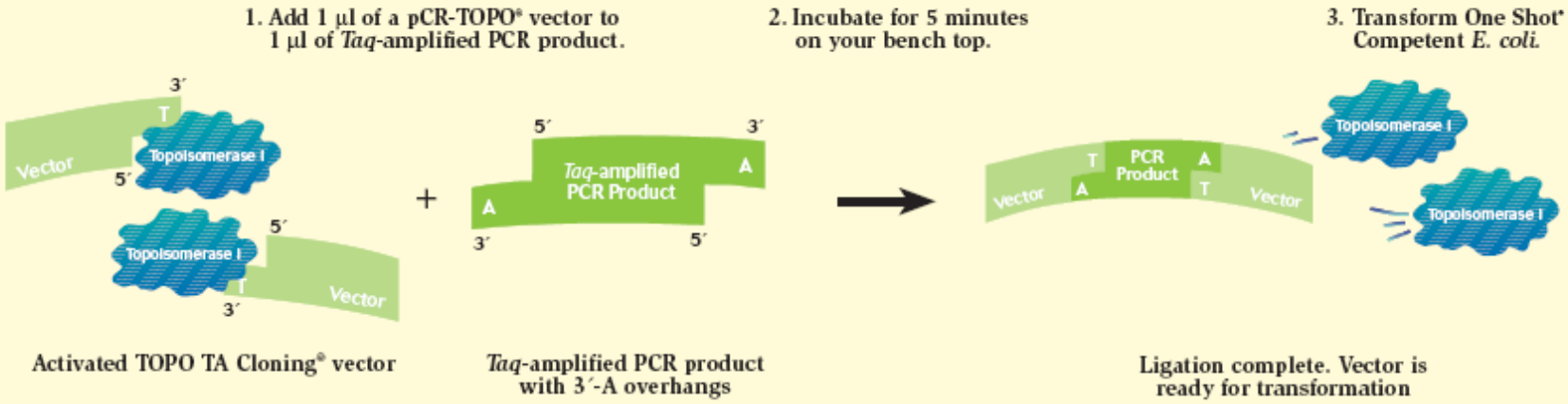
Digestion directe du produit PCR peut être inefficace: clonage du produit PCR dans vecteur TA…

Méthode de clonage: PCR

Direct ligation with TA Cloning® Technology

The TA Cloning® technology makes it possible to easily clone PCR products produced by *Taq* polymerase. *Taq* has a terminal transferase activity that adds a single 3´-A overhang to each end of the PCR product. TOPO TA Cloning® vectors contain 3´-T overhangs that enable the direct ligation of *Taq*-amplified PCR products (Figure 6)(2,3).

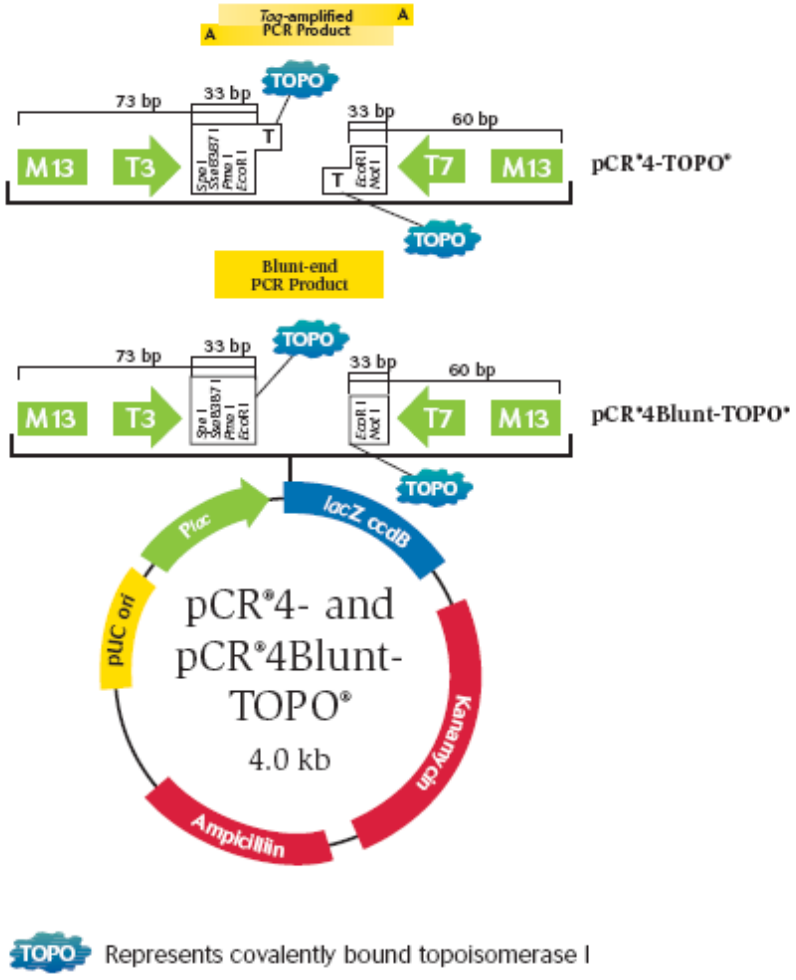
Figure 6 - How TOPO TA Cloning® works



Clonage en vecteur TA (TOPO ou autre…) : rapide et efficace

Digestion du vecteur de clonage PCR puis sous-clonage en vecteur d'expression

Méthode de clonage: PCR

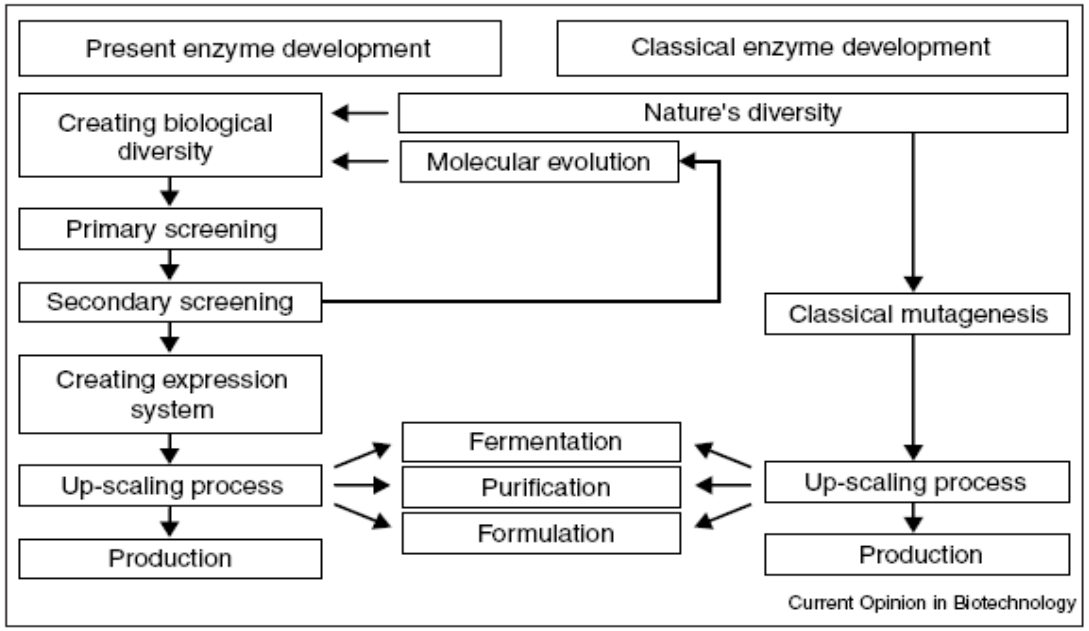


Vecteur TA : nécessite des 3'A (*Taq* polymérase)

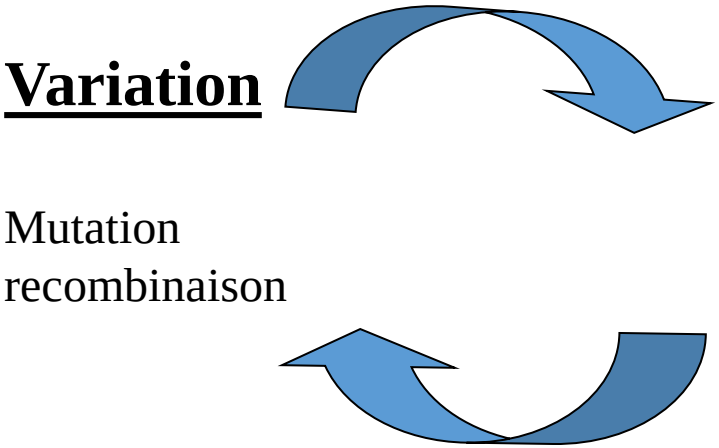
Vecteur « blunt »: proofreading polymerase (type *Pfu*)

Evolution dirigée des protéines

The steps involved in classical versus state-of-the-art development of enzymes.



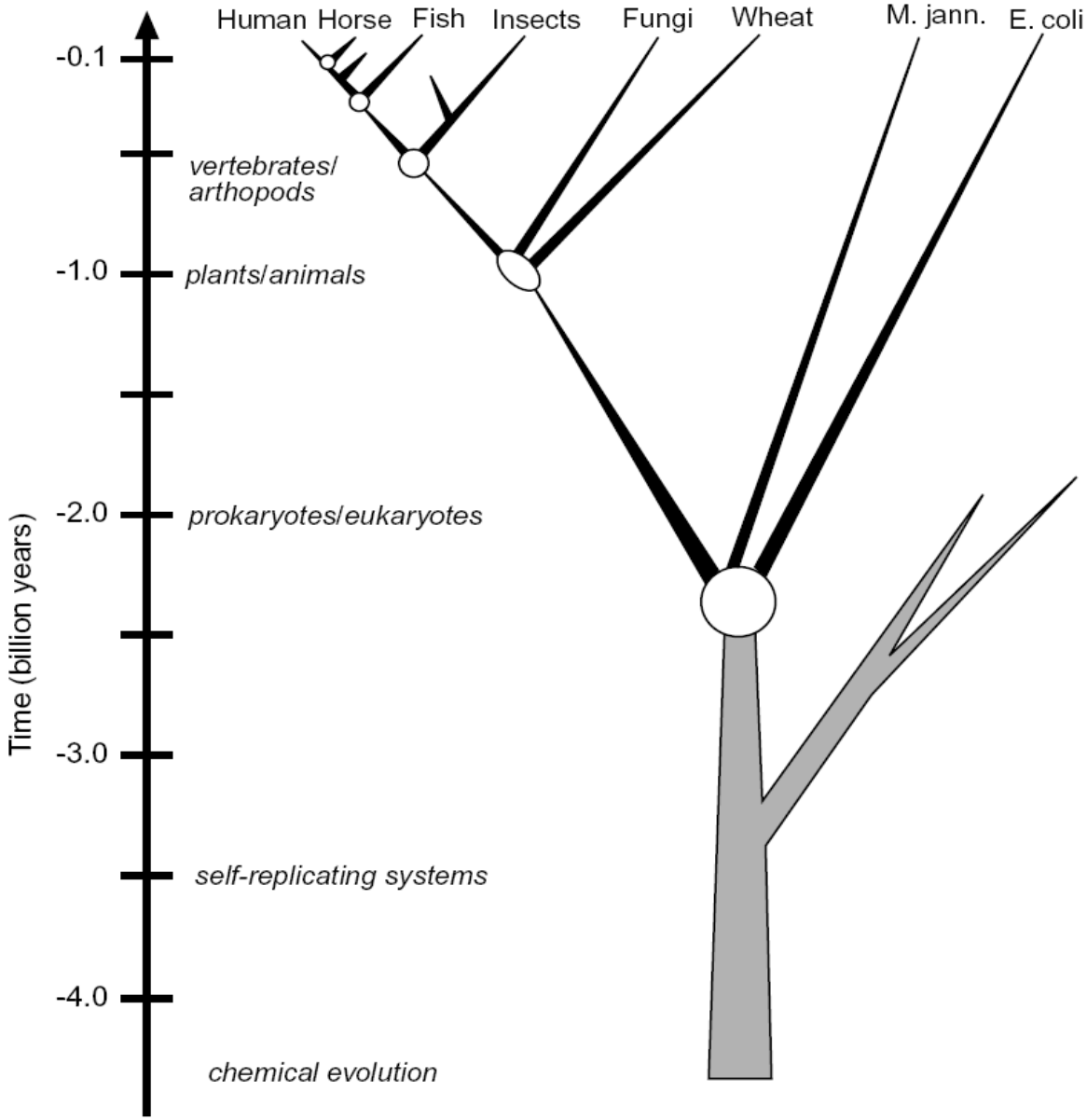
Évolution (cf Darwin) = sélection s'appliquant sur une variation dans la population



1 cycle = 1 génération

Adaptation au milieu
Réponses aux contraintes environnementales

Figure 1: The tree of life

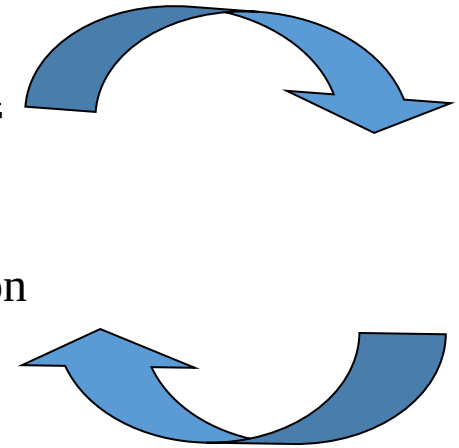


Adapted from Dayhoff *et al.*, 1978.

Apparition de nouvelles structures et fonctions

Application aux protéines : mêmes mécanismes mais environnement contrôlé

Conditions contrôlées
= contraintes environnementales choisies



Variation

Mutation
recombinaison

Sélection

Contrainte fonctionnelle:
Fixation, activité enzymatique,
solubilité, stabilité…

« Niveau » réglable

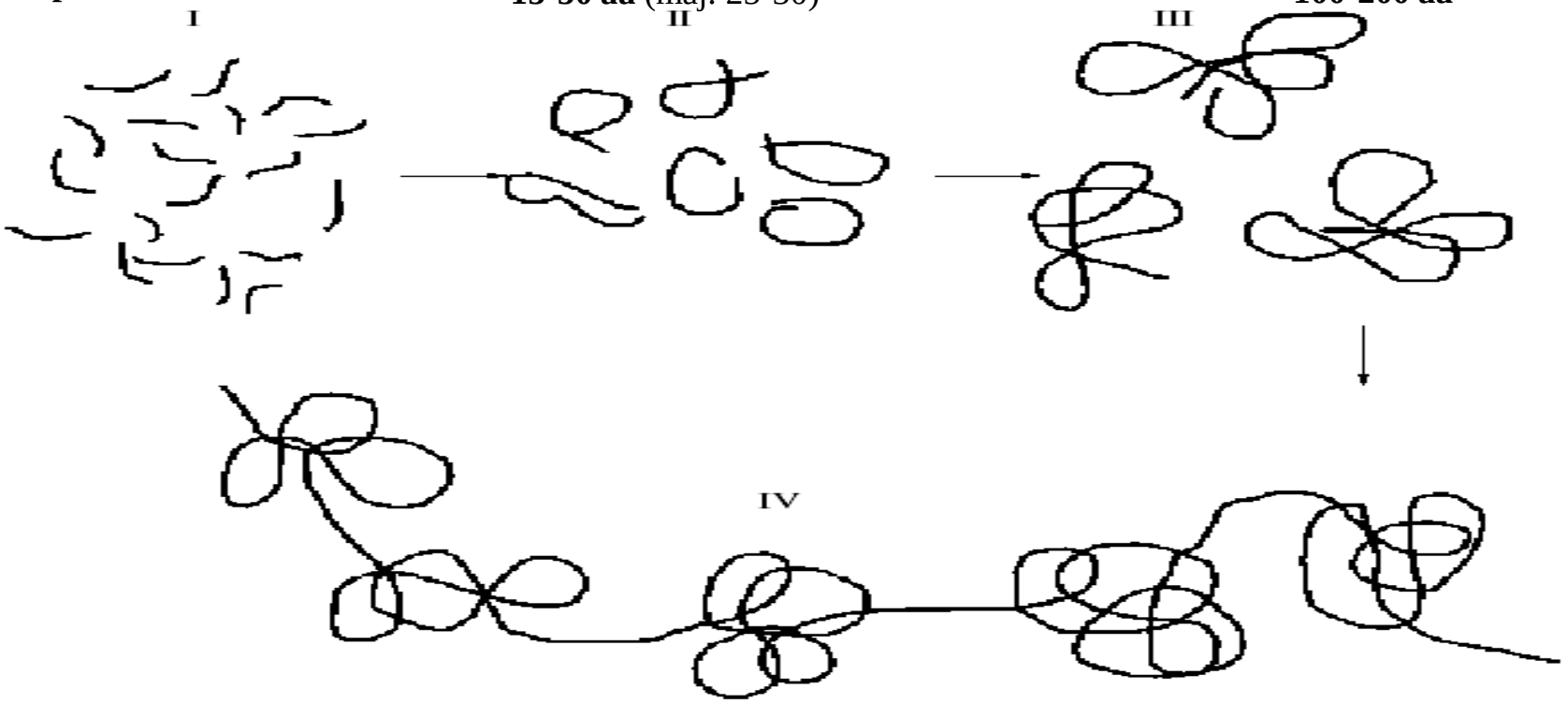
1 cycle = 1 génération

Notions d'évolution des protéines

Peptides: 5-10 aa

« boucles fermées »
15-50 aa (maj. 25-30)

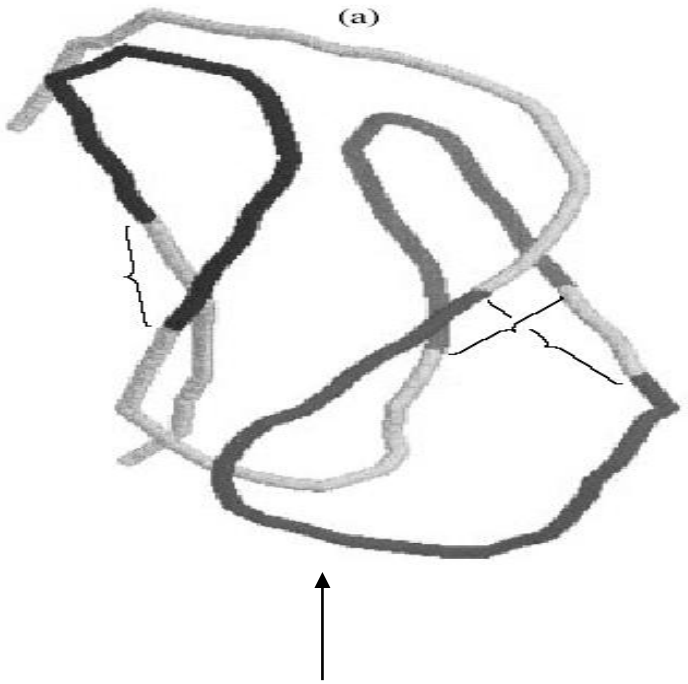
Domaines:
100-200 aa



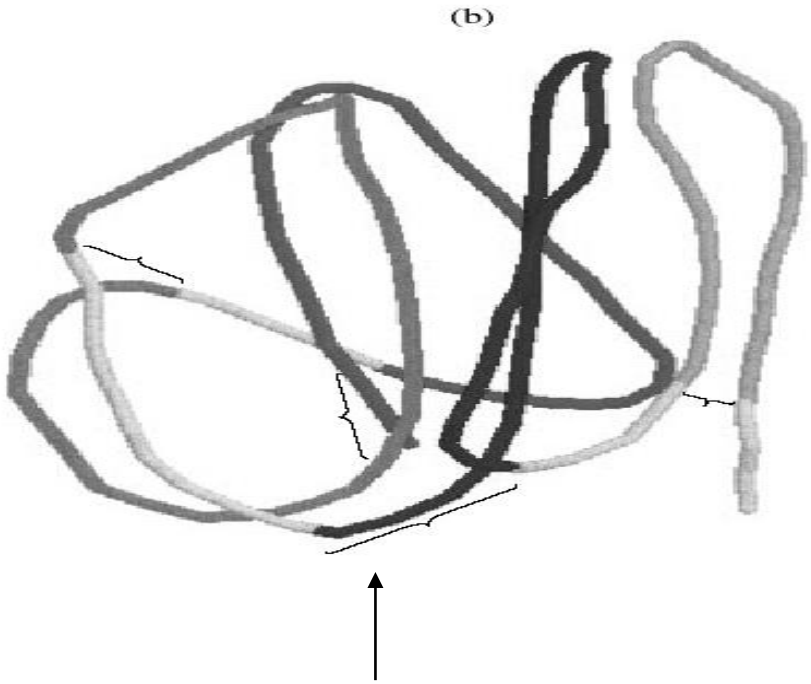
Protéines mulidomaines

Notions d'évolution des protéines

le stade domaine (« fold »)



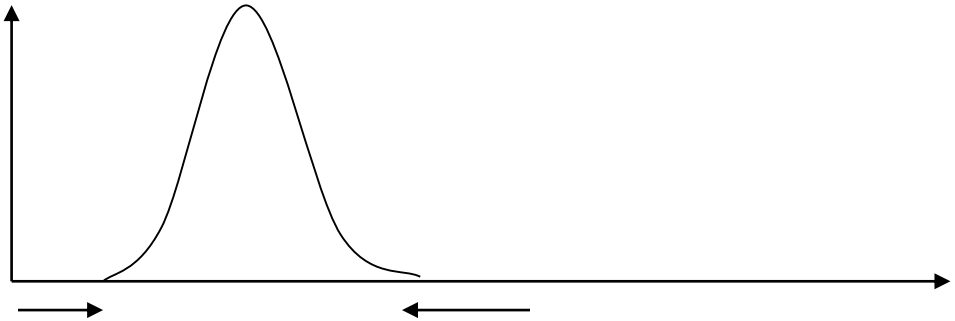
Globine: 3 boucles
141 résidus



flavodoxine: 4 boucles
138 résidus

Notions d'évolution des protéines

contraintes d'hier et d'aujourd'hui

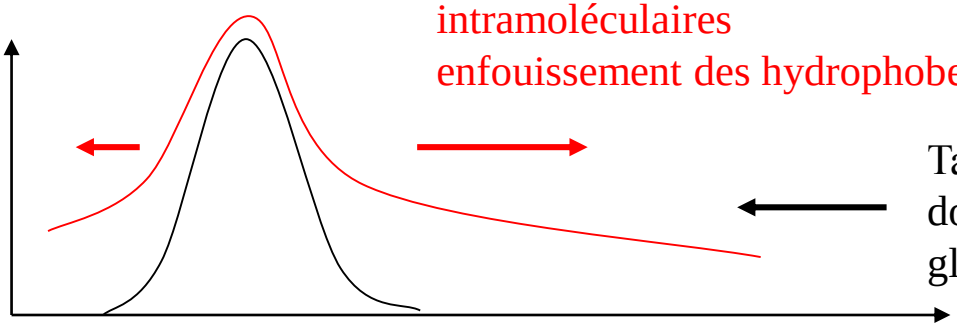


Stabilité ADN

Mecanisme de réplication
thermostabilité/forme active
solubilité

Séquence d'ADN codante au sein
d'une plus grande molécule

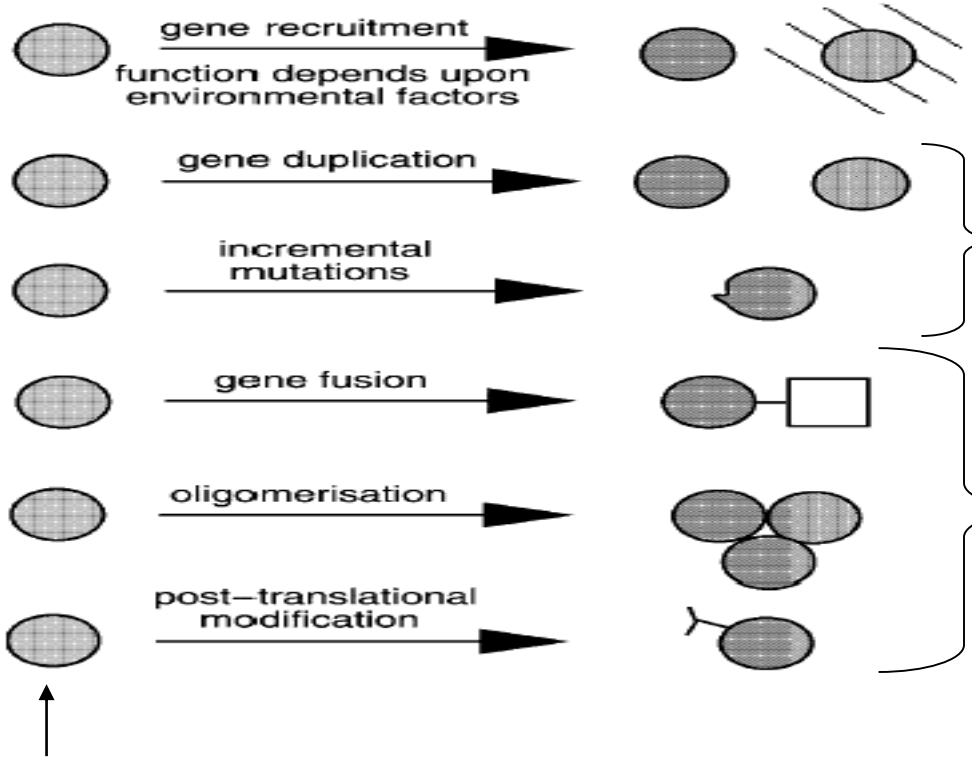
Stabilité apportée par les contacts
intramoléculaires
enfouissement des hydrophobes



Taille maximale d'un
domaine protéique
globulaire ~40 kDa (400 aa)

Notions d'évolution des protéines

Acquisition de nouvelles fonctions



Paramètres physico-chimiques modifiés

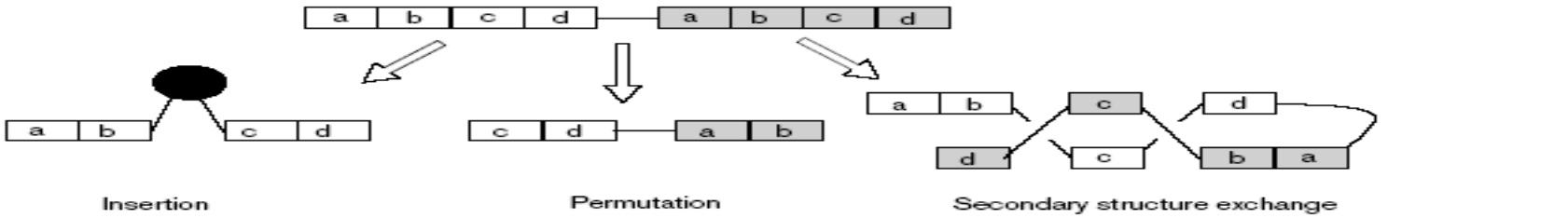
Séquence modifiée structure globale conservée

Modification du contexte protéique du module

Module de base

Notions d'évolution des protéines

Acquisition de nouvelles fonctions



Current Opinion in Structural Biology

Génération rapide de diversité

Contraintes fonctionnelles

Contraintes antigéniques

« rational design » vs « directed evolution »

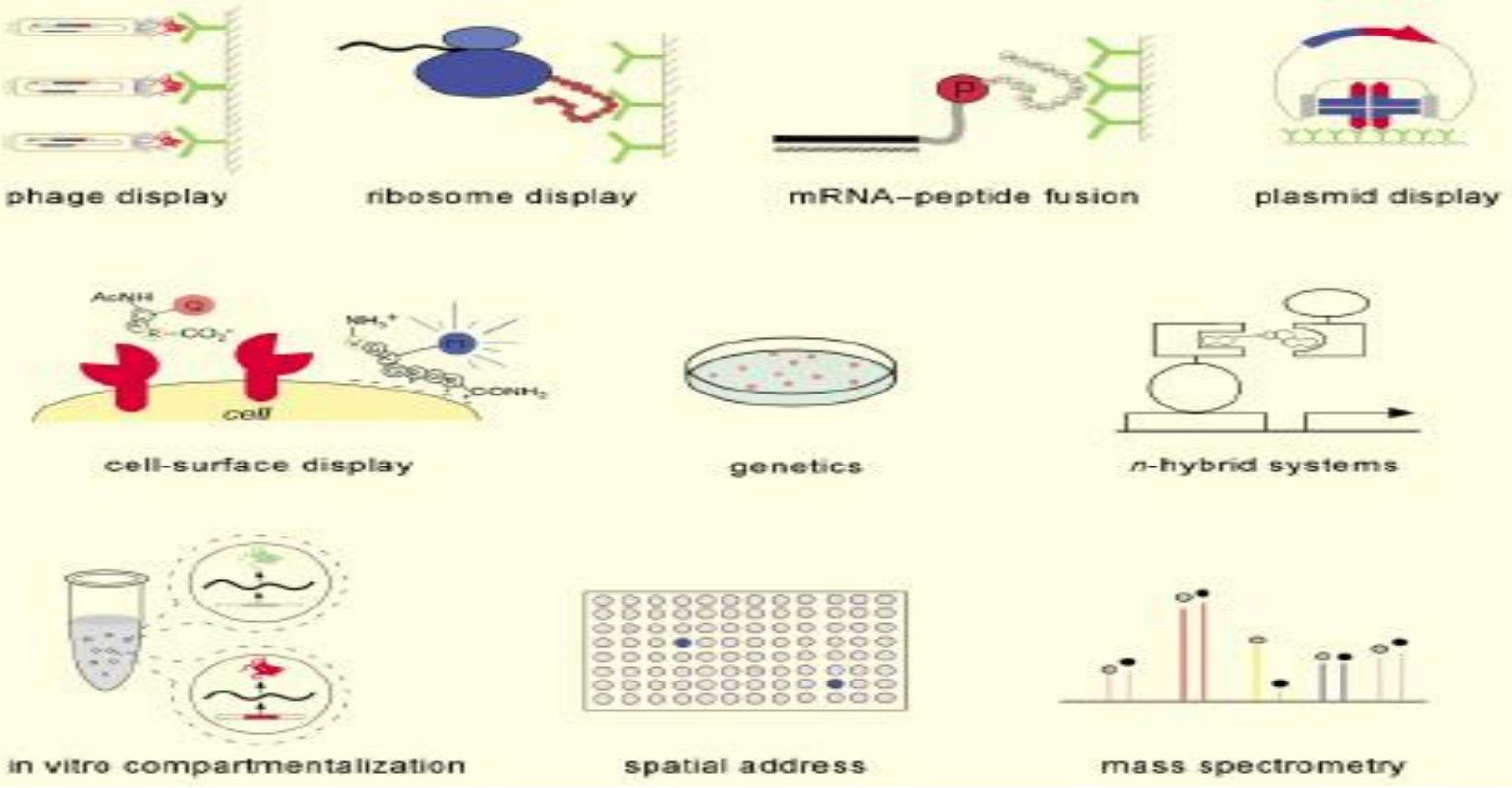
Modification rationnelle	Evolution dirigée
Connaissance fine de la relation structure-fonction	Connaissance d'un mécanisme de base (site actif, interface, ab initio?…)
Prédiction de l'effet de la modification	Sélection des meilleurs variants
Exploration d'un petit nombre de possibilité	Population diversifiée

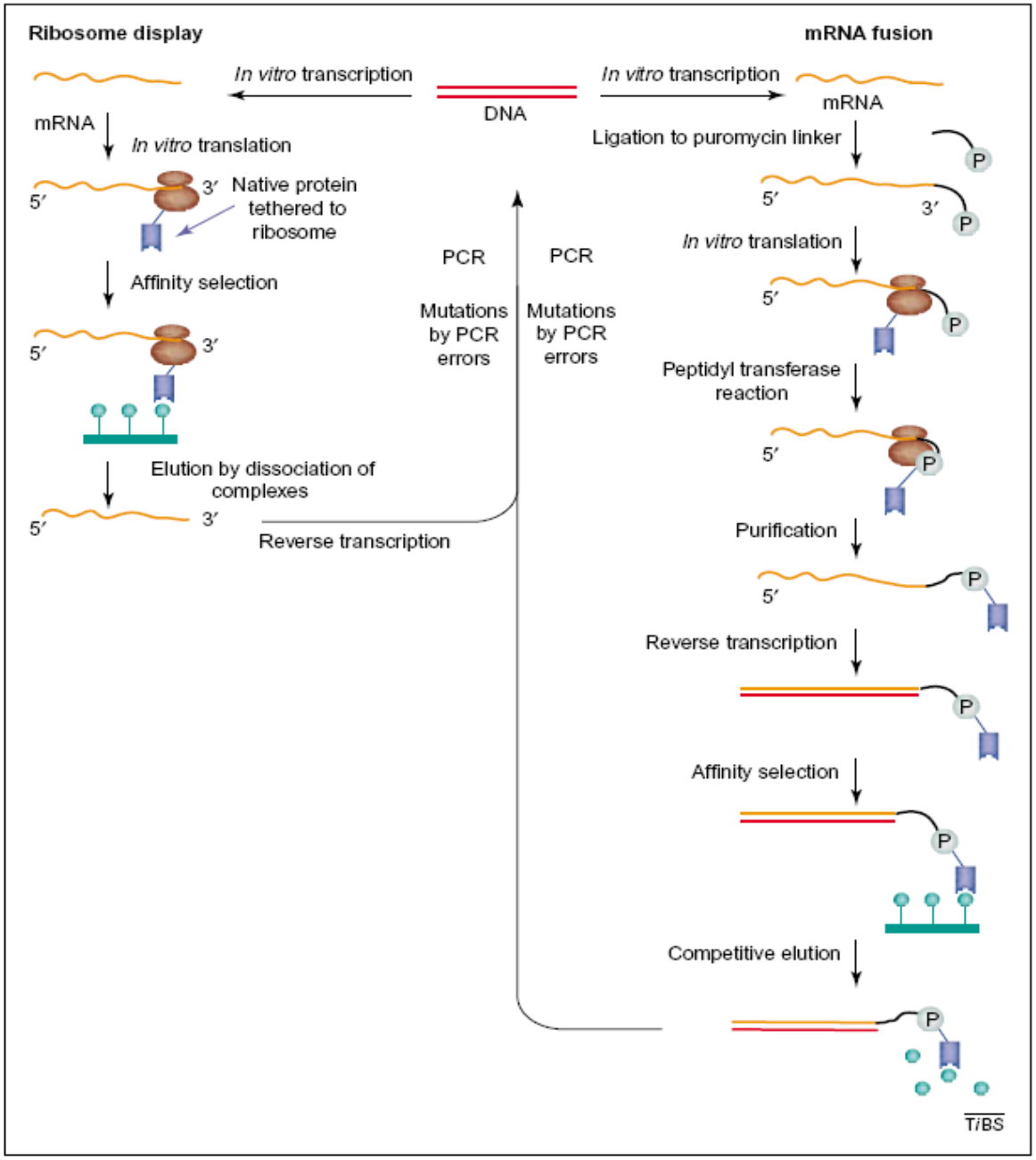
Liaison phénotype-information génétique

Sélection sur le phénotype (activité enzymatique, ligand…)
Utilisation de l'information génétique pour les applications biotechnologiques

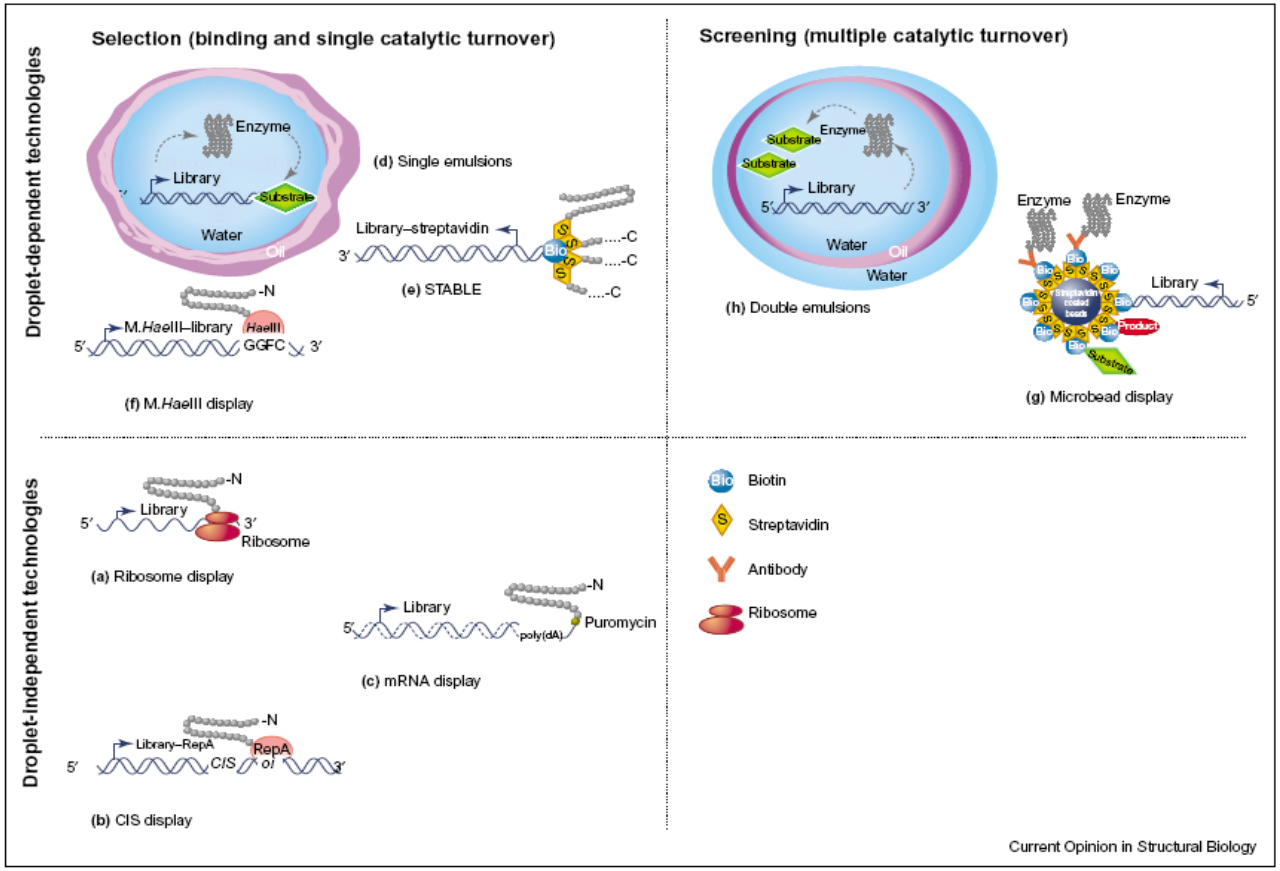
Cellule (clonable)	**Phage, plasmide**	***In vitro***
Surface display (bactérie, eucaryote)	Phage display plasmide display	Ribosome display mRNA display
Activité intracellulaire (antibiorésistance…)		Compartimentation (émulsion)

Cf: production in vitro des protéines


phage display
ribosome display
mRNA–peptide fusion
plasmid display
AcNH
NH3+
CONH2
cell
cell-surface display
genetics
n-hybrid systems
in vitro compartmentalization
spatial address
mass spectrometry


Ribosome display
mRNA fusion
In vitro transcription
DNA
In vitro transcription
mRNA
mRNA
In vitro translation
Ligation to puromycin linker
5′
3′
Native protein tethered to ribosome
P
In vitro translation
PCR
PCR
Affinity selection
Mutations by PCR errors
Mutations by PCR errors
Peptidyl transferase reaction
Elution by dissociation of complexes
Purification
Reverse transcription
Reverse transcription
Affinity selection
Competitive elution
TiBS

Liaison phénotype-information génétique



Overview of different *in vitro* genotype–phenotype linkages. Selections for binding are possible in all systems. Selections for catalysis can be subdivided into single turnover selections (mainly useful for an initial yes/no identification of catalysts beyond a fixed threshold) and multiple turnover selections (to differentiate catalysts quantitatively according to how much product has been produced). Single turnover catalysis is detected by association of the genotype–phenotype entity with a transition state analogue or a suicide inhibitor. Selections for multiple catalytic turnovers are possible in droplet-based systems in which the reaction product can be directly detected. The product concentration can be measured by FACS, either in double emulsions (h) or, after becoming attached to the bead, using a fluorescently labelled anti-product antibody (g). Droplet-independent technologies. These systems rely on a specific, essentially intramolecular reaction between nucleic acid and encoded protein; this obviates the need to keep them together in a droplet. **(a)** Ribosome display. The nascent protein does not dissociate from the ribosome and remains non-covalently associated with its coding mRNA template through the ribosome. **(b)** CIS display. A protein fusion to the DNA-binding protein RepA is directed by the *CIS* element to bind specifically to the *ori* site on the DNA template that it is expressed from. **(c)** mRNA display. The protein is covalently linked to its mRNA template through a puromycin linker that is covalently attached to the protein by ribosome-catalysed peptide bond formation. Droplet-dependent technologies. Protein expression occurs in droplets containing one copy of DNA, which ensures that each *in vitro* expressed protein is linked to the specific DNA that encodes it. **(d)** Simple emulsions. A droplet holds one gene encoding a protein that acts on a substrate connected to or identical with the DNA. **(e)** STABLE display. A protein fusion to streptavidin is linked non-covalently to its encoding biotin-labelled DNA. **(f)** M.*Hae*III display. A protein fusion to the DNA methyltransferase M.*Hae*III forms a covalent bond to a suicide inhibitor (5-fluorodeoxycytidine) that is integrated into the M.*Hae*III recognition sequence (GGFC) on its coding DNA template. **(g)** Microbead display. Polypeptide, substrate or product are linked to their template DNA via a series of non-covalent interactions based on streptavidin-coated beads. **(h)** Double emulsions. A droplet holds one gene encoding a protein that acts on free substrate.

Méthodes in vitro

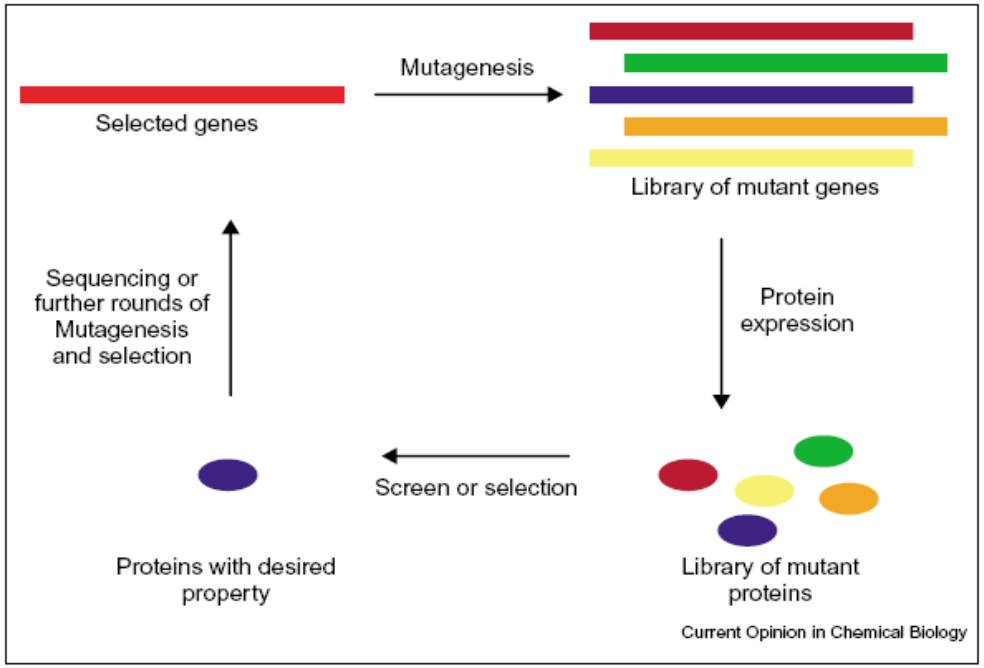
Liaison phénotype-information génétique

Table 1

Overview of the different methods used to link genotype and phenotype in directed evolution.

System	Affinity		Catalysis		Library size (per ml)
	Peptides	Proteins	Single turnover	Multiple turnover	
In vivo					
Cells: selection	[56]	–	–	[2]	Normally $<10^8$
Cells: screening	[57]	[58]	–	[2]	Normally $<10^6$
Phage display	[5]	[5]	e.g. [53,59]	–	$<10^{12}$ [6]
***In vitro* RNA display**					
mRNA display	[27,31–33,35]	[29,30,34,36,37••]	Possible	–	$<10^{14}$
Ribosome display	[14,17]	[7,15,16,18,19, 21–23,24••]	[60]	–	$<10^{12}$
***In vitro* DNA display**					
STABLE (non-covalent DNA display)	[44,46]	[45]	Possible	–	$<10^{10}$
CIS (non-covalent DNA display)	[25]	Possible	Possible	–	$\sim10^{12}$
Covalent DNA display (M.*Hae*III fusion)	[49]	[49]	Possible	–	$<10^{10}$
Emulsion droplets	–	–	[39]	[40,41•,55•]	$<10^{10}$
Microbeads in droplets	[52]	[61]	Possible	[51••]	$<10^{10}$

Nature cyclique des méthodes d'évolution



Population de départ diversifiée (+/- : mAb, peptide, banque naïve…)

Diversification à chaque cycle (mutation neutre ou négative)

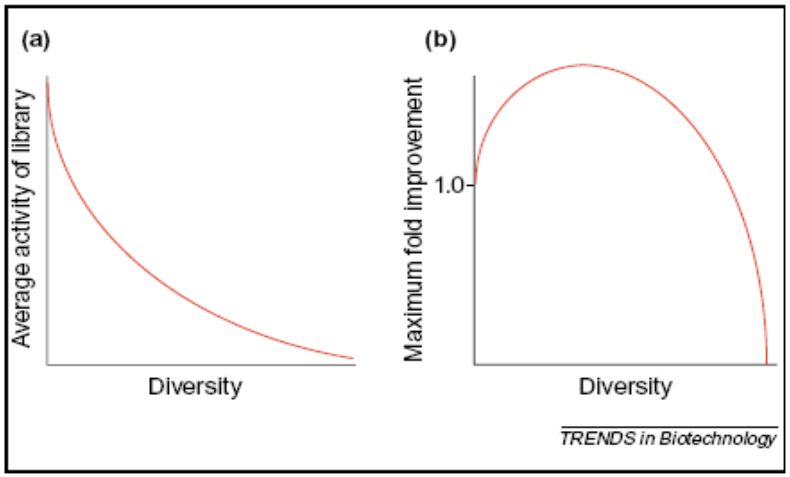
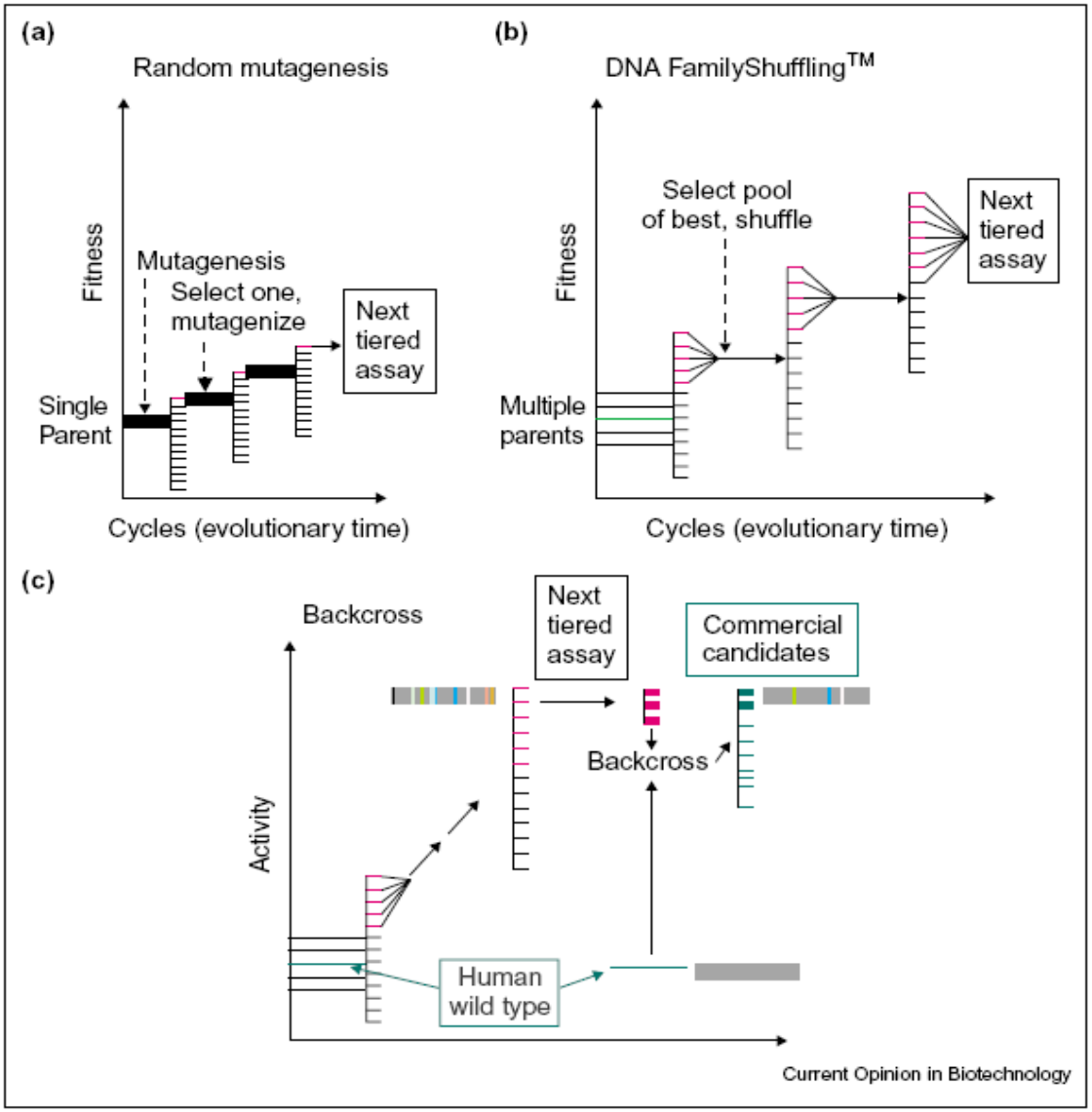
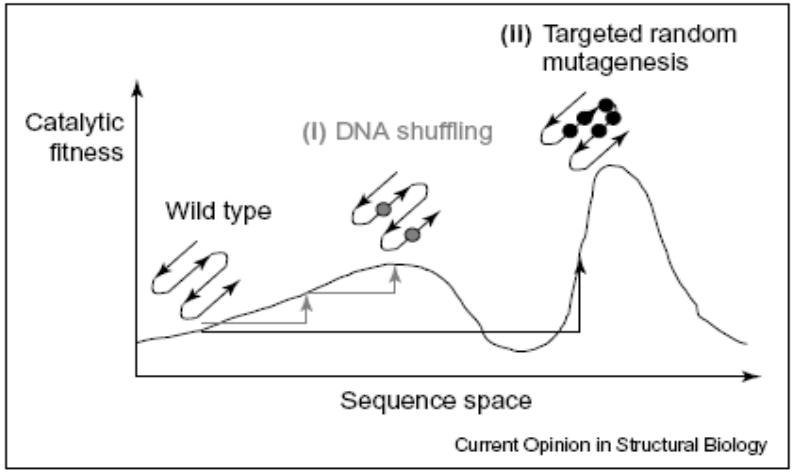


Fig. 2. Postulated relationships between diversity in a library and (a) average activity of library and (b) improvement of best variant. The shape of (b) is as suggested from a statistical mechanical model of the relationship between maximum improvement and average DNA mutation rate of a single gene [17].

Nature cyclique des méthodes d'évolution



Nature cyclique des méthodes d'évolution



A simplified evolutionary fitness landscape representing peaks of enzyme catalysis in sequence space. **(i)** Incremental improvement by repeated error-prone PCR or DNA shuffling can find the local catalytic fitness maximum. **(ii)** Targeted mutation of relevant enzyme regions, such as the active site, may access new peaks or functions in the fitness landscape by identifying multiple synergistic mutations.

Existence de «puits » : nécessité d'exploration large

Génération de la diversité et méthodes de sélection

Liaison à un ligand (Ab, recepteur, peptide…) : panning

Activité enzymatique : intracellulaire, émulsion ou phage

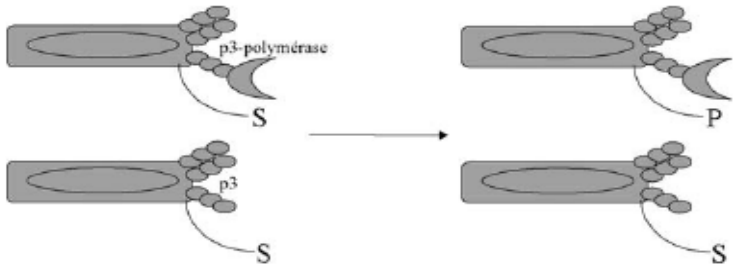


Fig. 1. Principe de la sélection en fonction de l'activité catalytique. Le substrat S de l'enzyme est lié de manière covalente à une protéine de la surface du phage. Si une enzyme est exprimée à la surface du phage à proximité du substrat, le substrat est converti en produit P, qui est isolé par chromatographie d'affinité ; la phage-enzyme est ainsi extraite du mélange de phages-protéines. Si aucune enzyme n'est exprimée à la surface du phage, le substrat S n'est pas converti en produit P et le phage n'est pas isolé par chromatographie d'affinité pour le produit de la réaction.

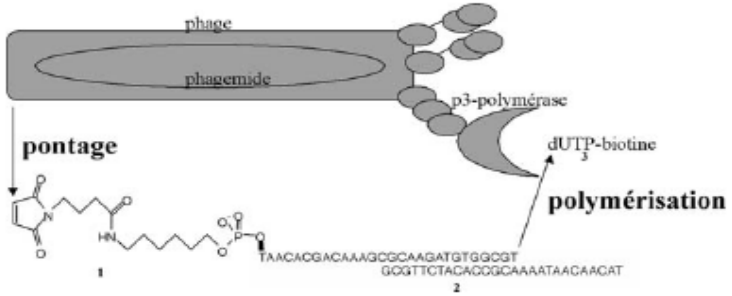
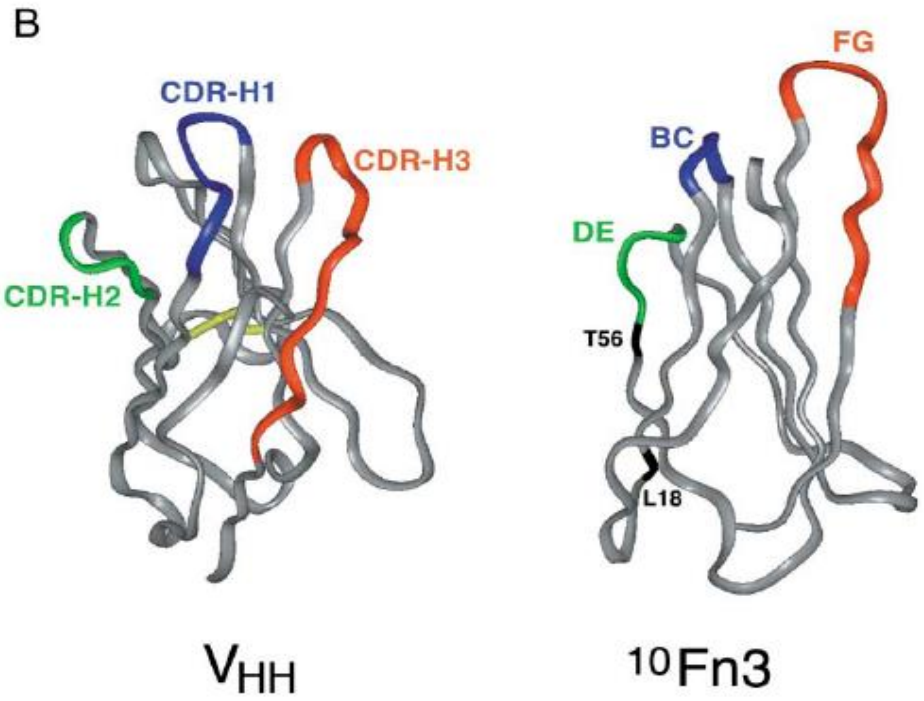


Fig. 2. Sélection in vitro en fonction de l'activité ADN-polymérase. La réaction de pontage conduit à la formation d'une liaison covalente entre l'amorce **1** modifiée par un groupement maléimidyle à son extrémité 5' et un nucléophile à la surface du phage. La réaction de polymérisation est l'addition catalysée par la polymérase d'un nucléotide **3** à l'extrémité 3' de l'amorce **1** hybridée au brin matrice **2**. Le trait en gras entre O et T indique l'oxygène 5' du nucléotide T de l'amorce.

Génération de la diversité et méthodes de sélection



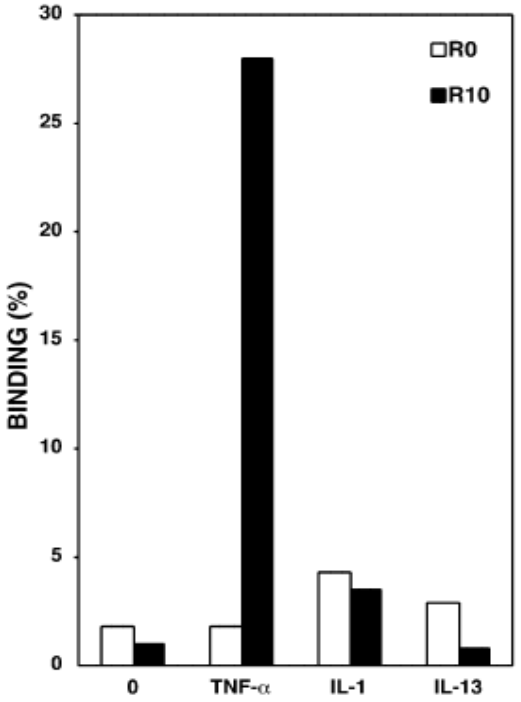
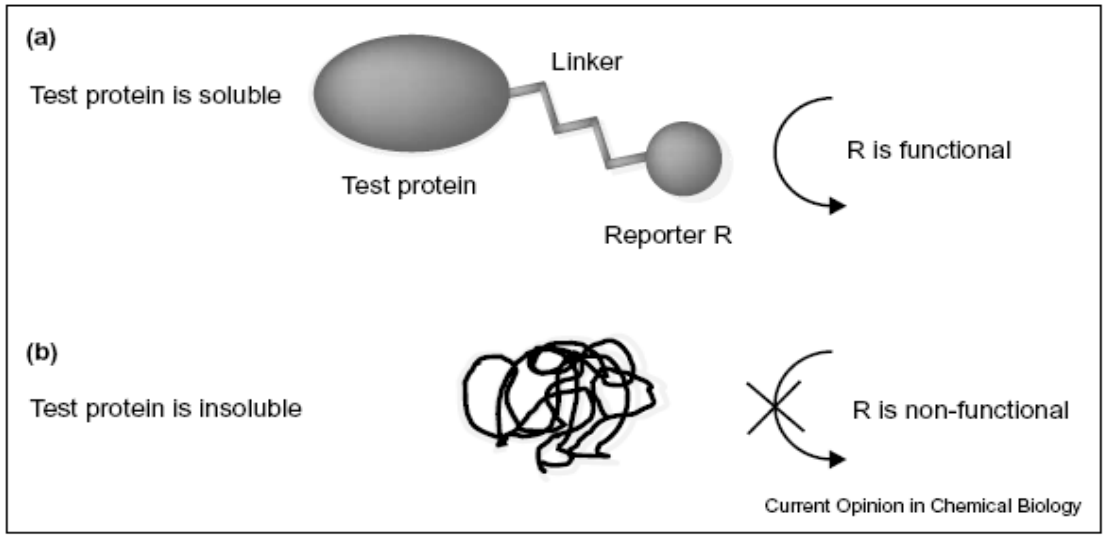


Figure 2. Specificity of Binding of the ^{10}Fn3 Pool Translated from the Original, Randomized Library and from the Pool Obtained after Ten Rounds of Selection for Binding to TNF-α

The library or the PCR product of the elution after R10 was transcribed and translated in vitro, in the presence of ^{35}S-methionine but without forming RNA-protein fusion. The fraction of the resulting free protein bound to underivatized Sepharose, to TNF-α-Sepharose (~10 μM TNF-α in the target-^{10}Fn3 binding mixture), to IL-1α-Sepharose (~30 μM IL-1α), and to IL-13-Sepharose (~50 μM IL-13) is compared. The R10 selected protein pool binds specifically to TNF-α-Sepharose, whereas the R0 library binds at a comparable, low background level to all three immobilized cytokines.

Point de départ de la banque : domaine varianble d'un anticorps ou domaine d'adhésion (ex: fibronectine)

Génération de la diversité et méthodes de sélection



Principle of fusion solubility reporter. A reporter domain with an easily measured function is fused to a test domain via a flexible linker. **(a)** When the test protein is soluble, the reporter domain is functional and its activity can be detected. **(b)** On the other hand, when the test protein is insoluble or misfolded, the function of the reporter domain is compromised. The reporter domain can be any moiety with an easily measured function. The mode of the assay depends only on the function of the reporter, which can be a selectable marker protein such as a metabolic enzyme or antibiotic resistance protein or a screenable marker such as an enzyme or fluorescent protein. Intact proteins or fragments of proteins can be used, such as split protein interactors.

Optimisation de la solubilité d'une protéine

Génération de la diversité et méthodes de sélection

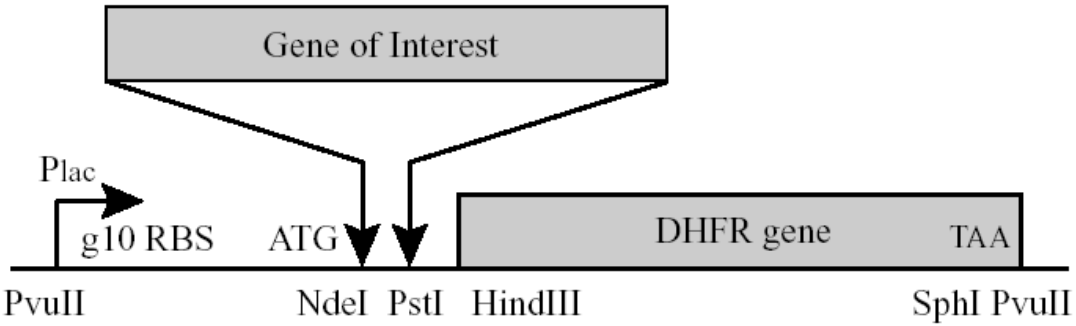
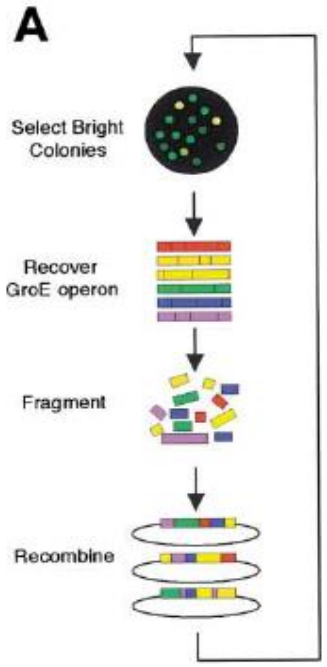
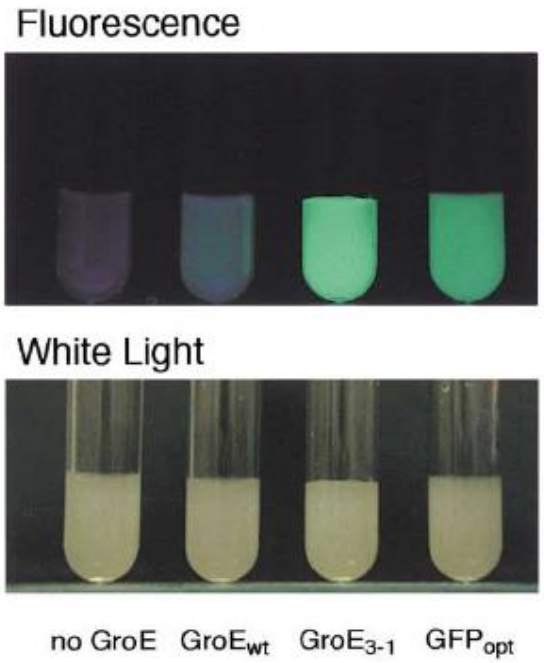
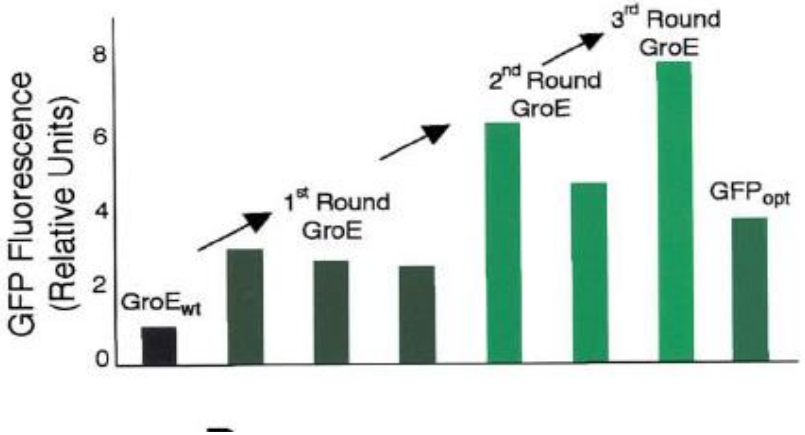


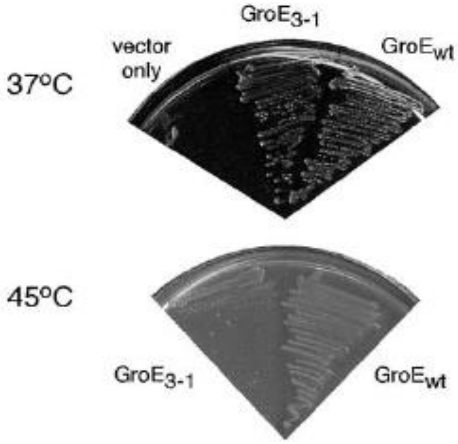
Fig. 1. The *Escherichia coli* DHFR fusion vector pJWL1030folA. The important elements in the cloning region are shown.

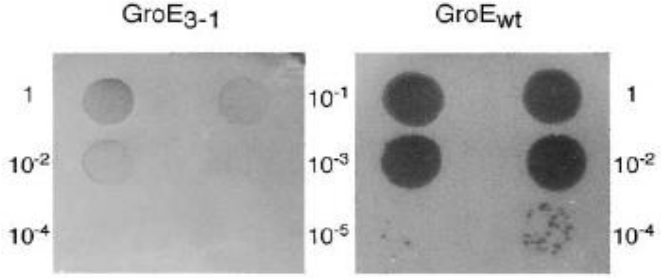






Sélection sur une propriété peu contre-sélectionner une autre





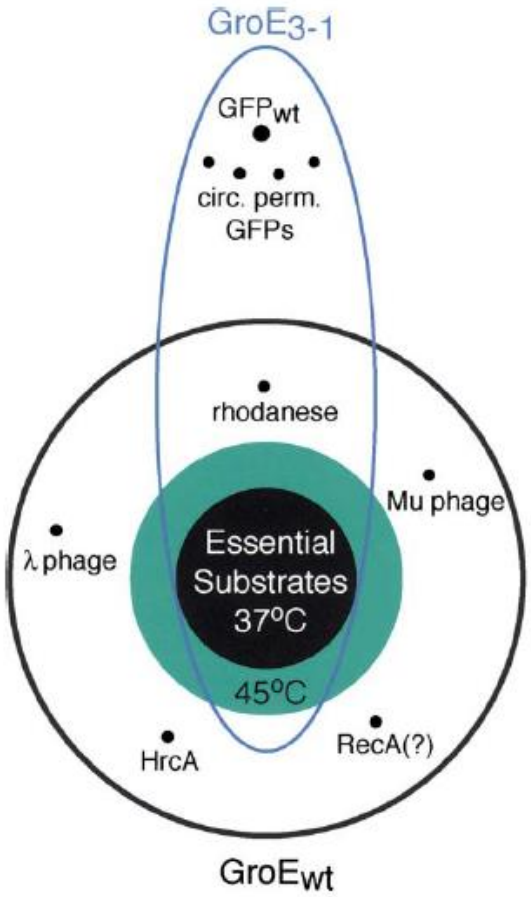


Figure 7. Venn Diagram Illustrating the Shift in the Substrate Range of $GroE_{3-1}$

Indicated are the sets of proteins optimally folded by $GroE_{wt}$ (black circle) and $GroE_{3-1}$ (blue oval). The solid black disc represents substrates needed for growth at 37°C and the pale green disk represents substrates important for growth at 45°C. The increased sensitivity to MMS might result from a decreased ability of $GroE_{3-1}$ to fold RecA (Houry et al., 1999). T4 phage growth is supported by both wild-type and 3-1 variants of GroEL but does not depend on the bacterial encoded GroES (Ang et al., 2000).

Génération de la diversité et méthodes de sélection

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Jan. 2005, p. 442–450 Vol. 71, No. 1
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.1.442–450.2005

Isolation of Llama Antibody Fragments for Prevention of Dandruff by Phage Display in Shampoo

Edward Dolk,[1] Marcel van der Vaart,[2] David Lutje Hulsik,[1] Gert Vriend,[3] Hans de Haard,[2]†
Silvia Spinelli,[4] Christian Cambillau,[4] Leon Frenken,[2] and Theo Verrips[1,2]*

Department of Molecular and Cellular Biology, University of Utrecht, Utrecht,[1] Unilever Research and Development Vlaardingen, Vlaardingen,[2] and CMBI, KUN, Nijmegen,[3] The Netherlands, and Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités Aix-Marseille I & II, Marseille, France[4]

exemple de conditions de sélection en fonction de l'application

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Biotechnology Advances 20 (2002) 229–238

BIOTECHNOLOGY ADVANCES

www.elsevier.com/locate/biotechadv

DNA shuffling: induced molecular breeding to produce new generation long-lasting vaccines

Sergio H. Marshall

Laboratorio de Genética e Immunología Molecular, Instituto de Biología, Facultad de Ciencias Básicas y Matemáticas, Universidad Católica de Valparaíso, Casilla 4059, Valparaiso, Chile

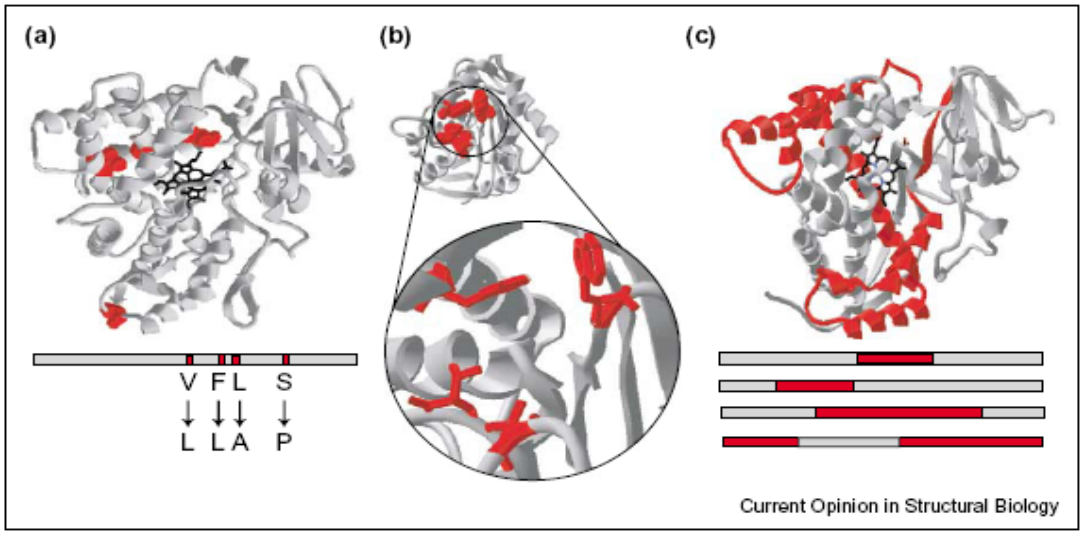
Accepted 15 April 2002

Évolution dirigée des protéines : réponse à la diversité antigénique?

Génération de la diversité et méthodes de sélection

Mutagenèse : error-prone PCR, souche mutante,…
recombinaison : DNA shuffling et variantes

Mécanismes similaires « dans la nature » (exon, boucles…)



Random mutagenesis, targeted mutagenesis and recombination are three strategies for producing sequence libraries for directed evolution. **(a)** Random mutagenesis introduces amino acid substitutions throughout the protein and can uncover beneficial mutations distant from the active site. The red residues in the structure at top show four mutations uncovered by random mutagenesis that enhanced the activity of mammalian cytochrome P450 2B1 on several substrates [15]. The bar below shows the distribution of mutations in the primary sequence. **(b)** Targeted mutagenesis uses structural information to select a small number of amino acids for randomization and is a good method for making multiple active site mutations. The red residues were targeted for mutagenesis to increase the enantioselectivity of *Pseudomonas fluorescens* esterase [26•]. **(c)** Recombination of protein sequences can introduce large numbers of simultaneous sequence changes. The structure shows a chimera of two bacterial cytochrome P450 sequences (fragments from different parents are shown in red and gray) that exhibits an altered functional spectrum [48]. Chimeras are created by recombining fragments of different protein sequences, indicated by the bars. The PDB structures in the figure are 1SUO, 1VA4 and 2HPD, respectively.

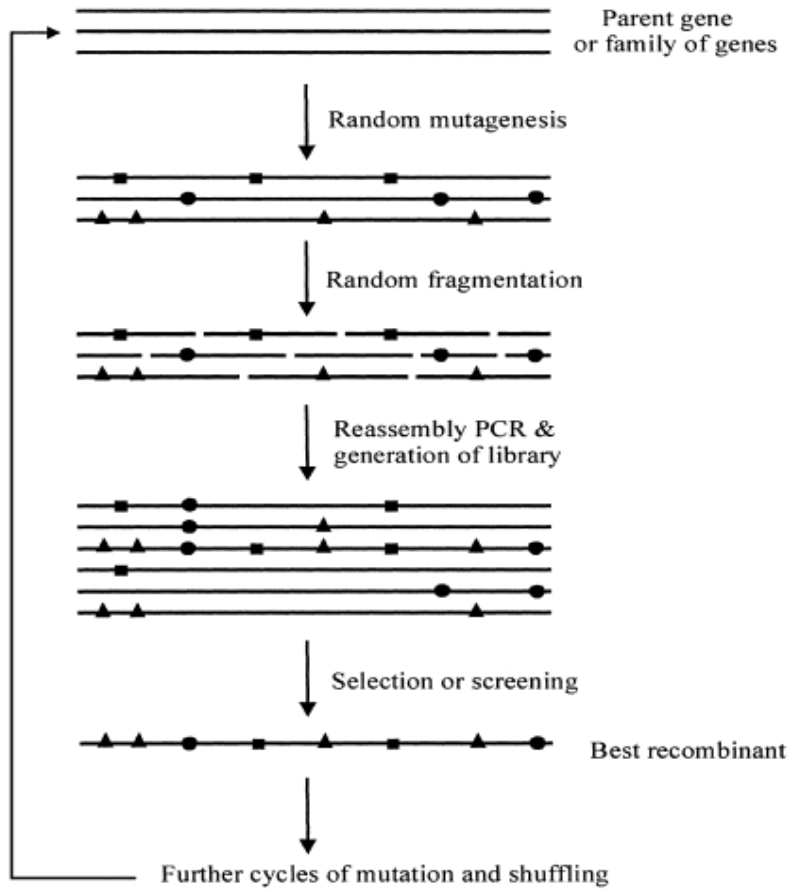


FIG. 5. Steps involved in the process of *in vitro* evolution by DNA shuffling. The parent gene or a pool of genes are mutagenized and randomly fragmented. Fragments containing different point mutations are reassembled in a PCR reaction, where the fragments act as primers for one another. Reassembled fragments are cloned in a suitable expression vector and a gene bank is generated. Finally, the mutant with desirable characteristics is selected by either screening or selection strategies and the mutant can be subjected to further cycles of mutation to accumulate beneficial mutations.

Combinaison de mutagenèse et recombinaison

Table 1

Summary of new technologies for generating molecular diversity.

Methodology	Target	Library size	Outcome	Advantage/disadvantage	Ref.
Oligonucleotide-directed randomization					
MAX randomization	Zinc-finger protein	1×10^2	More uniform codon distribution; bias from genetic code removed	Reduces target library size	[5•]
Codon shuffling (hexamer assembly)	TEM-1 β-lactamase	1×10^3	Severely truncated, divergent sequences retain near-WT activity	Randomization not constrained by length of parental gene	[7]
Whole gene randomization					
Sequence Saturation Mutagenesis(SeSaM)	Green fluorescent protein	nr	100 clones sequenced; 88% of point mutations attributable to SeSaM	Unbiased or controllable mutation spectrum	[8]
Homology-dependent recombination					
Degenerate homoduplex recombination (DHR)	Mammalian epidermal growth factors	3×10^6	Mouse–human hybrid has 123-fold greater mitogenic potency	All shuffled variants equally likely	[10••]
Synthetic shuffling	15 *Bacillus* subtilisins	1.5×10^3	20% of variants active; improved pH and thermostability profiles	All shuffled variants equally likely	[11•]
Assembly of designed oligonucleotides (ADO)	*B. subtilis* lipases LipA and LipB	3×10^3	Twofold improvements in conversion and/or enantioselectivity	All shuffled variants equally likely	[12•]
Recombination-dependent exponential amplification PCR (RDA-PCR)	Green and yellow fluorescent proteins	nr	All clones analyzed have fluorescence similar to parents	Guarantees odd number of crossovers	[14]
Mutagenic and unidirectional reassembly (MURA)	*Serratia* phospholipase PlaA	3×10^3	Lipase activity increased 1000-fold; no decrease in phospholipase activity	Combines epPCR, DNA shuffling and unidirectional random truncation	[15]
Recombined extension on truncated templates (RETT)	Two *Serratia* chitinases	8×10^2	1.5-fold increase in activity; epPCR improved thermostability	Removes sequence bias of DNase I digestion	[16]
Homology-independent recombination					
Sequence-independent site-directed chimeragenesis (SISDC)	TEM-1 and PSE-4β-lactamases	1.5×10^3	Of 256 possible variants, 14 retain up to 30% WT activity	Number and locations of crossovers specified	[20•]
Structure-based combinatorial protein engineering (SCOPE)	DNA polymerases β and X	6.5×10^5	Hybrids with up to five crossovers complement auxotroph	Some variability in linkers between specified subdomain building blocks	[21]
Enhanced crossover SCRATCHY	Rat and humanGSTs	1×10^5	Average of 1.4 crossovers per gene; 1% retain parental activity	Increases number of crossovers; some variability in crossover location	[24•]

Abbreviations: WT, wild-type; GST, glutathione-*S*-tranferase; nr, not reported.

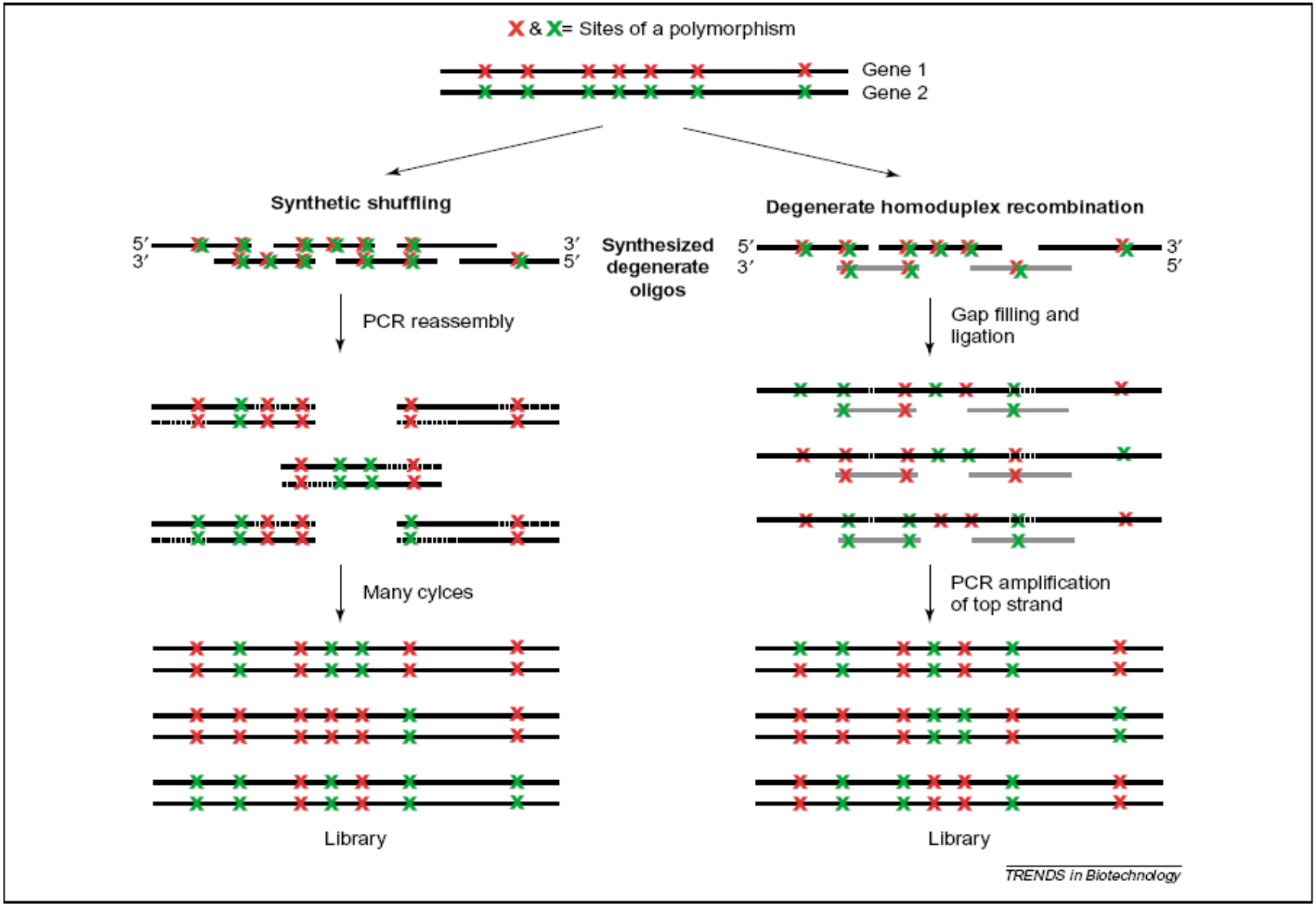


Fig. 1. Schematic representation of synthetic shuffling and degenerate homoduplex recombination. For simplicity, recombination of only two genes with only seven polymorphisms is shown. In synthetic shuffling, the gene is divided into equal length, overlapping, degenerate oligonucleotides that are reassembled in a PCR reaction. For degenerate homoduplex recombination, the oligonucleotides are designed to minimize divergence in the overlap region and the genes are reassembled on a series of scaffold oligonucleotides.

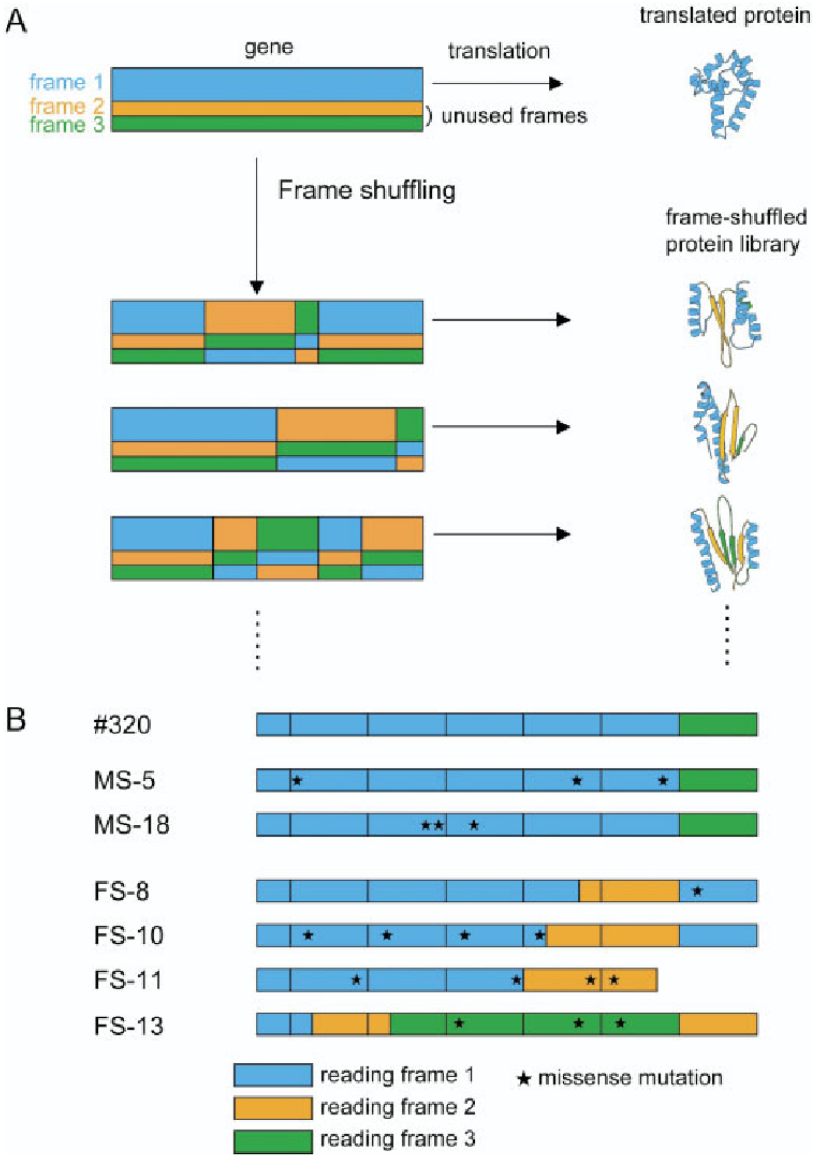


Fig. 1. Concept of frame shuffling and proteins used in this study. (A) Frame shuffling mutagenesis; (B) schematic structures of the artificial proteins used.

« frame shuffling »

Génération de la diversité et méthodes de sélection

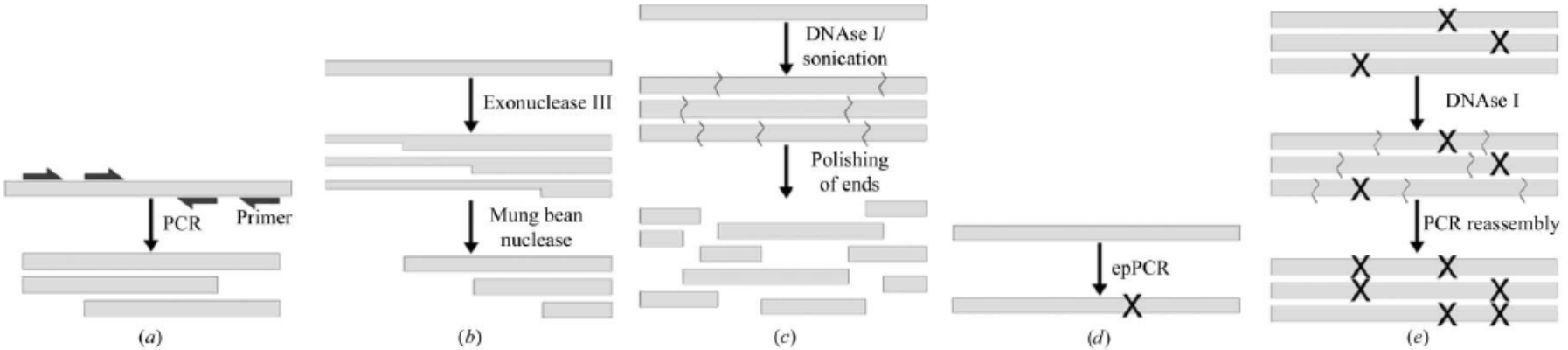
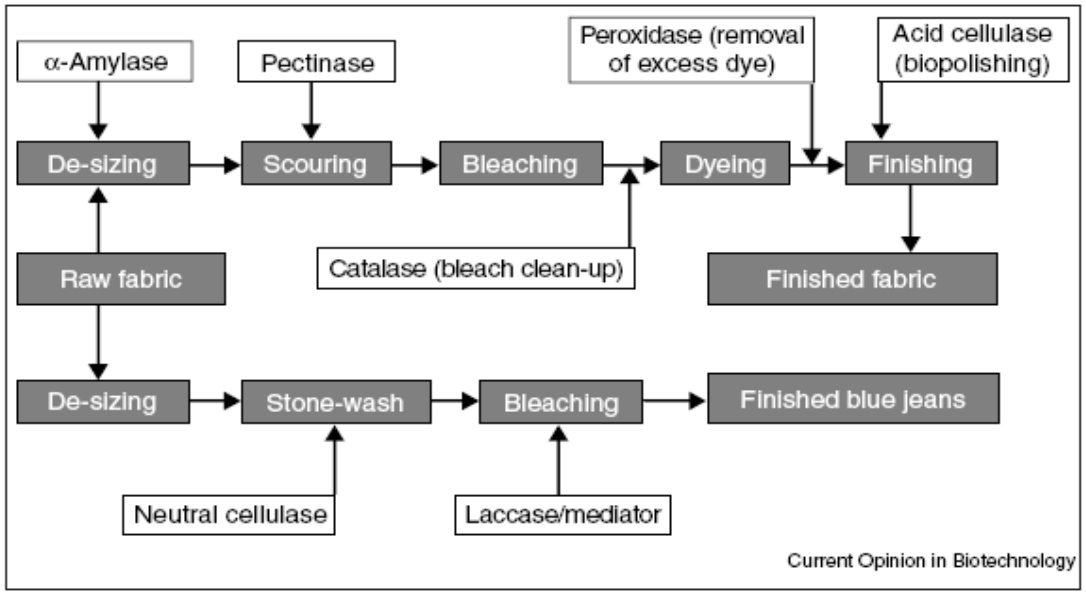


Figure 2
DNA truncation and mutagenesis methods. (*a*) Generation of specific constructs by PCR, (*b*) unidirectional DNA truncation using exonuclease III and nuclease (mung bean or S1), (*c*) fragmentation using enzymatic or physical breakage of the DNA, (*d*) random point mutation by error-prone PCR (epPCR) or bacterial mutator strain and (*e*) DNA shuffling for combining different lineages of point mutations on the same DNA strand.

Table 3

Modified recombinant protein pharmaceuticals approved in the USA.

Product name and reference	Product	Compared with endogenous	First indication	Reason for change
Betaseron [49] (Chiron)	IFNβ-1b	Cys17→Ser, glycosylation, no Met at N terminus	Multiple sclerosis	Increased production yields
CEA-Scan* [50] (Immunomedics)	Anti-carcinoembryonic antigen Ab fragment	^{99m}Tc-labeled	Detection of colorectal cancer	Imaging
Cerezyme [51,52] (Genzyme)	Glucocerebrosidase	Arg495→His, glycosylation	Type I Gaucher's disease	Change introduced during cloning, cell targeting
Enbrel [53] (Immunex)	Soluble TNFR	Dimeric TNFR fused to IgG1 Fc	Rheumatoid arthritis	Enhanced affinity and half-life
Geref [54] (Serono)	Human GHRH (synthetic)	N terminus only	Idiopathic pediatric growth hormone deficiency	Minimal biologically active core
Herceptin [55] (Genentech/Roche)	Anti-HER2 Mab	Humanized Mab	Metastatic breast cancer	Reduced immunogenicity, enhanced affinity
Humalog [56] (Lilly)	Insulin	Reversal of Lys and Pro at C terminus of B chain	Diabetes	Faster onset of action
Infergen [57] (Amgen)	Type 1 IFNα	Consensus sequence of four α subtypes, four additional amino acid changes	Hepatitis C	Increased activity, facilitated molecular construction
Lantus [58] (Aventis)	Long-acting insulin	Asp21→Gly and addition of two Arg to C terminus of B chain	Diabetes	Increased duration of action
Leukine [59] (Immunex)	Human GM-CSF	Pro23→Leu, glycosylation	Bone marrow transplantation	Expression/production
Miacalcin† (Novartis)	Calcitonin–salmon (synthetic)	50% identical to human	Postmenopausal osteoporosis	50 Times more potent than human version
Mylotarg [60] (AHP/Celltech)	Anti-CD33 Mab	Humanized, fused calicheamicin	Acute myeloid leukemia	Reduced immunogenicity, target toxin
Neumega (Genetics Institute)	Human interleukin-11	Lacks N-terminal Pro	Thrombocytopenia (chemo-induced)	Expression
Neupogen† (Amgen)	Human G-CSF	Additional N-terminal Met, nonglycosylated	Neutropenia (chemo-induced)	Expression
OncoScint* [61]	Anti-TAG-72 Mab	^{111}In-labeled	Detection of metastases of	Imaging



Enzymes used in various unit operations in textile wet processing and the manufacturing of Denim.