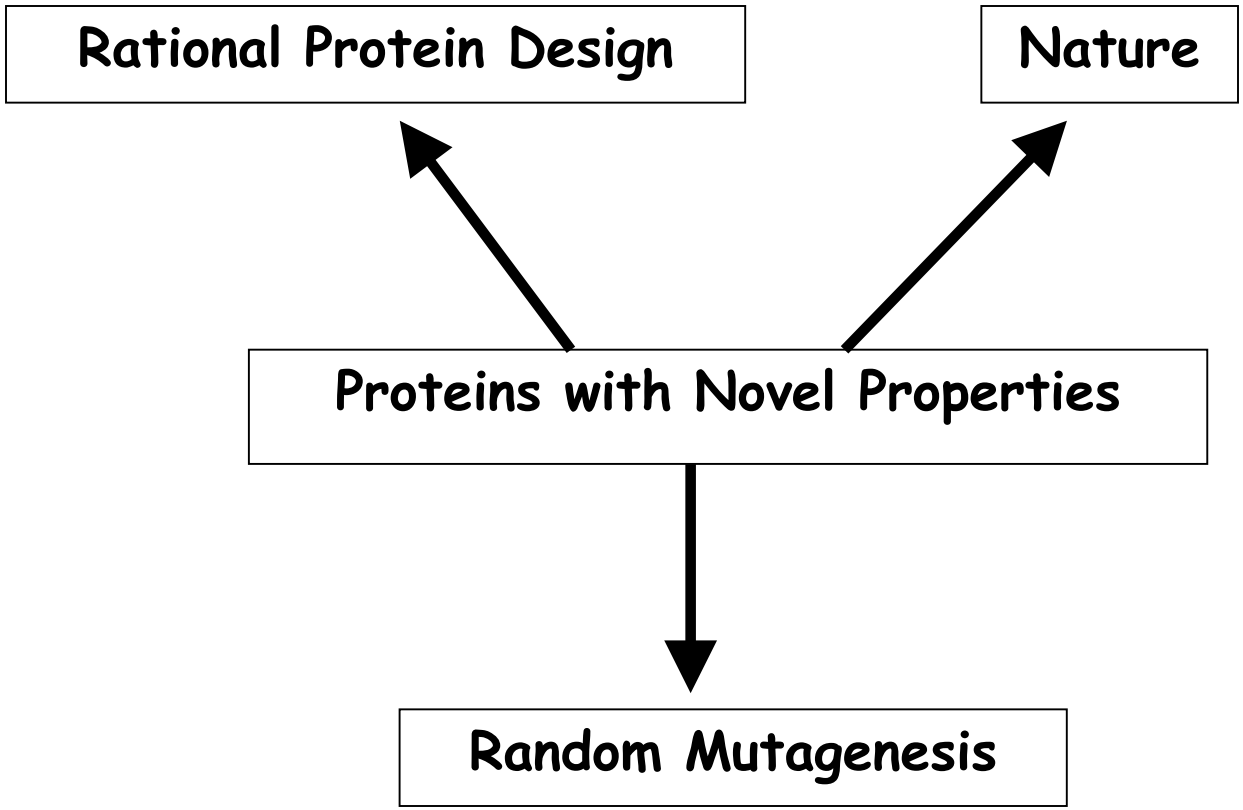
PROTEIN ENGINEERING

Protein engineering-Why?

- Enhance stability/function under new conditions
  - temperature, pH, organic/aqueous solvent, [salt]
- Alter enzyme substrate specificity
- Enhance enzymatic rate
- Alter epitope binding properties

Protein Engineering

Obtain a protein with improved or new properties



Evolutionary Methods

- **Non-recombinative methods:**

  -> Oligonucleotide Directed Mutagenesis (saturation mutagenesis)

  -> Chemical Mutagenesis, Bacterial Mutator Strains

  -> Error-prone PCR

- **Recombinative methods** -> Mimic nature's recombination strategy

  Used for: Elimination of neutral and deleterious mutations

  -> DNA shuffling

  -> Invivo Recombination (Yeast)

  -> Random priming recombination, Staggered extention precess (StEP)

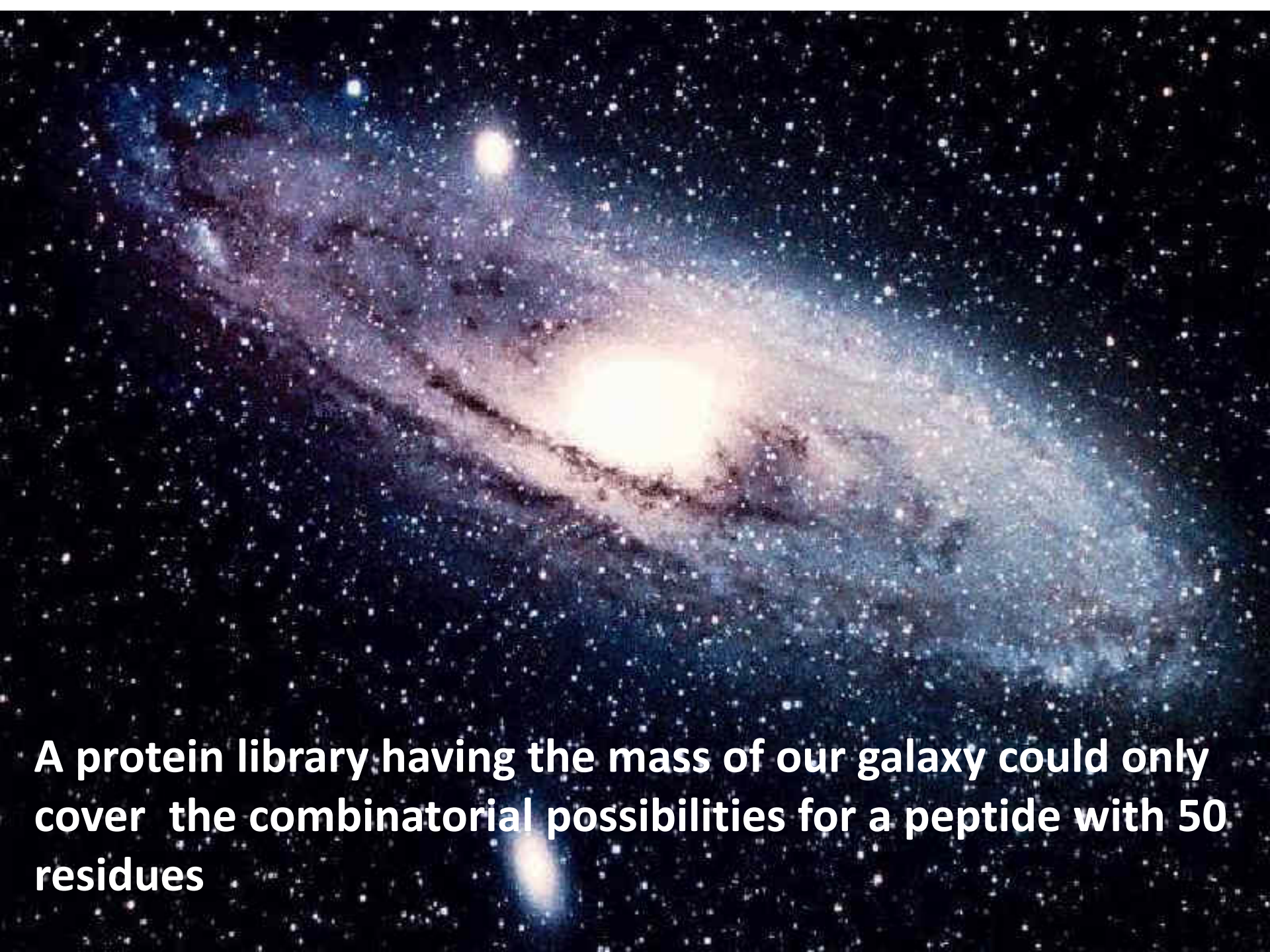
  -> ITCHY

RATIONAL DESIGN

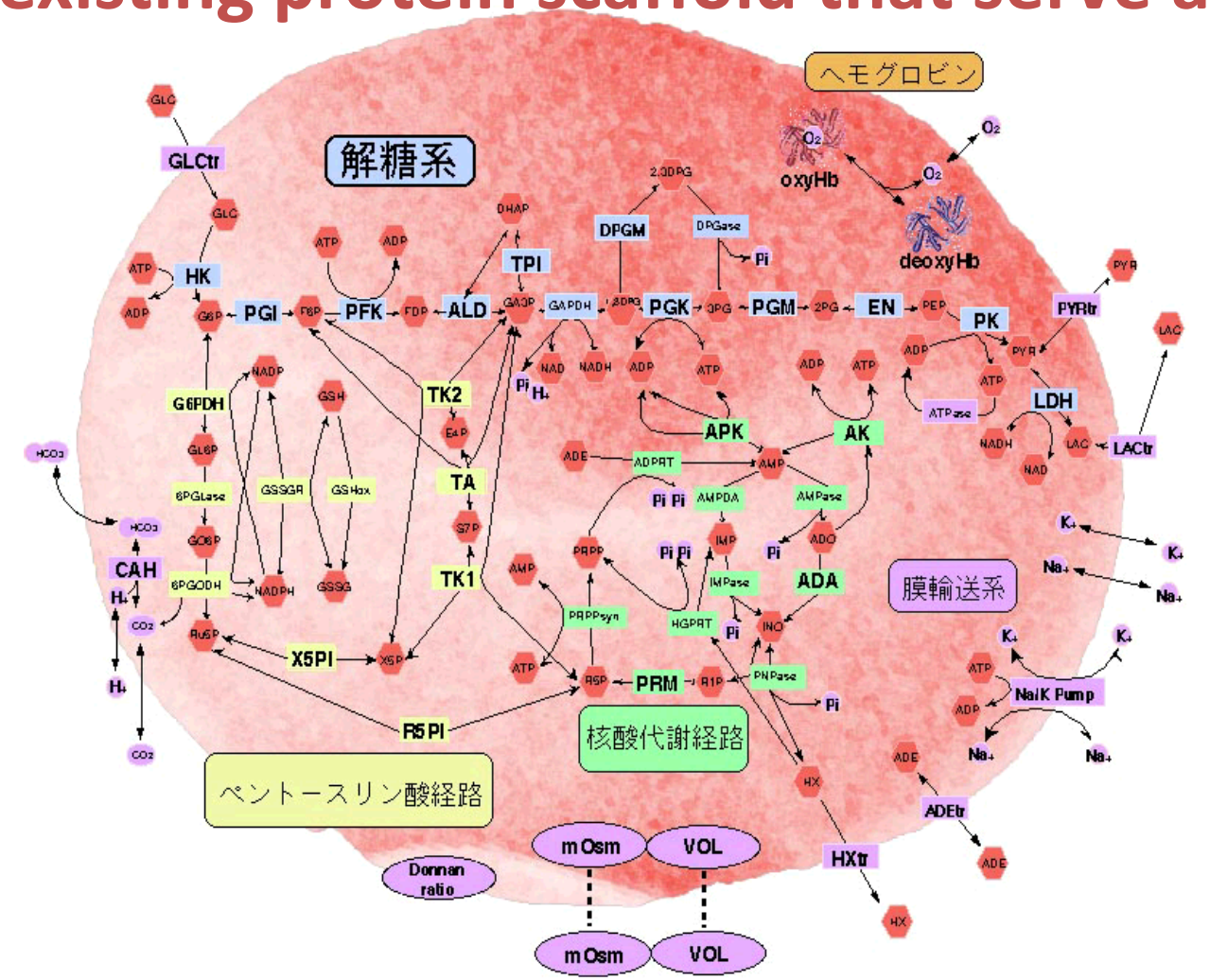
-Site directed mutagenesis of one or more residues

-Fusion of functional domains from different proteins to create chimaeric (Domain swapping)

Functional evaluation


A protein library having the mass of our galaxy could only cover the combinatorial possibilities for a peptide with 50 residues

Therefore even genetic selection approaches for designing novel functional proteins will not generally build on fully random sequences, but will be based on existing protein scaffold that serve as template.

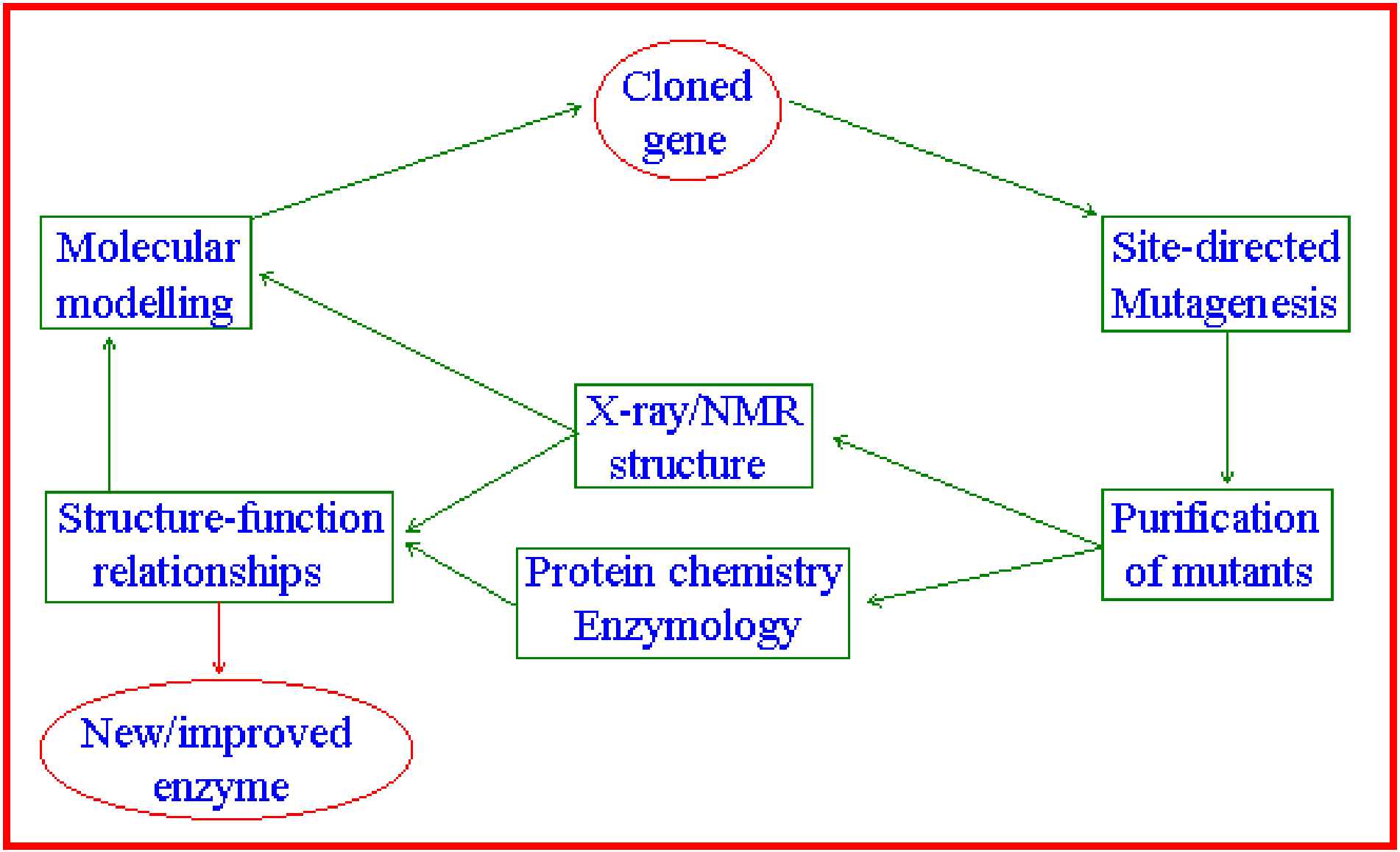


In order to consider the rational design of a target enzyme, one needs to have several pieces of information:

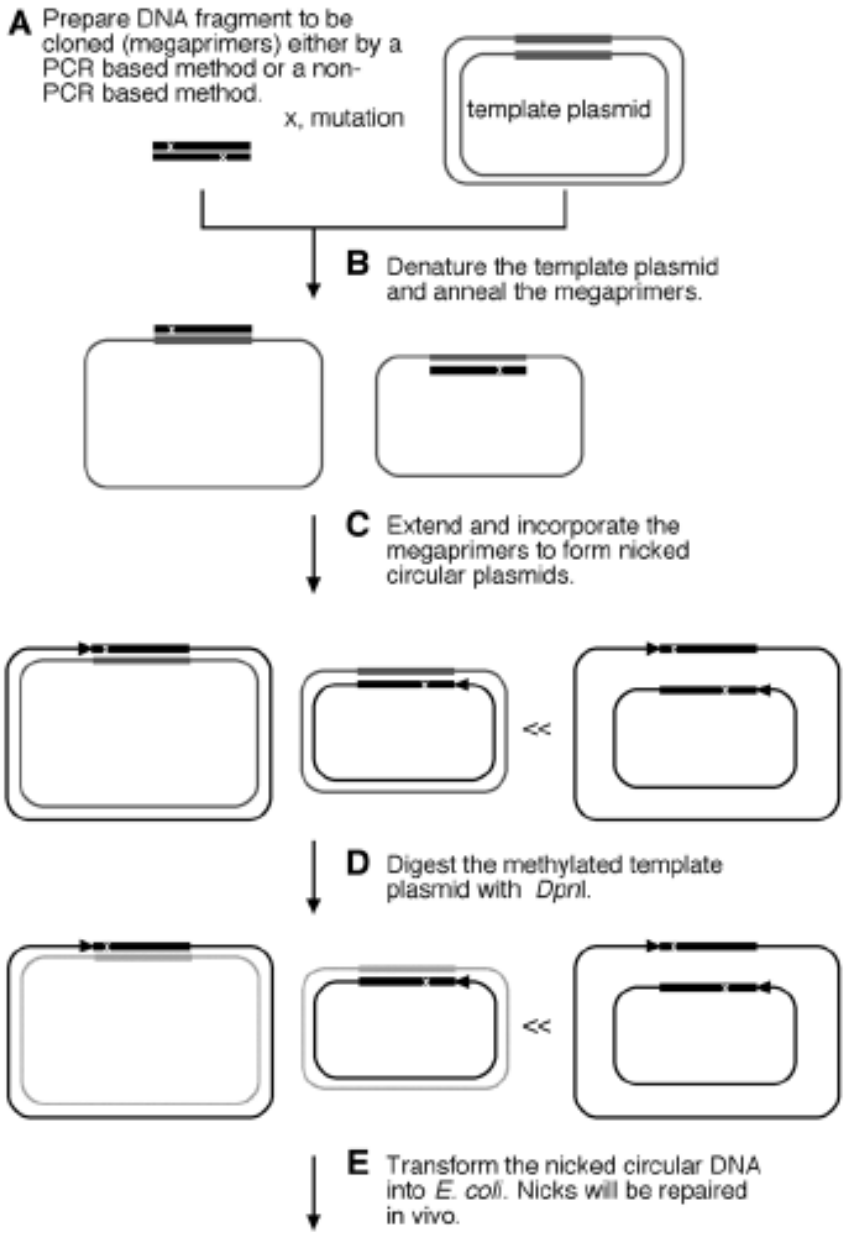
1. A cloned gene coding for the enzyme.
2. The sequence of the gene.
3. Information on the chemistry of the active site, ideally one would know which amino acids in the sequence are involved in activity.
4. Either a crystal/NMR structure for of the enzyme, or the structure of another protein displaying a high degree of structural homology.

The above information is needed in order to have a clear idea of which amino acids one should mutate to which likely effect.

Typically, protein engineering is a cyclic activity involving many scientists with different skills:



Creating Random Mutagenesis Libraries by Megaprimer PCR of Whole Plasmid (MEGAWHOP)



Construction of Designed Protein Libraries Using Gene Assembly Mutagenesis

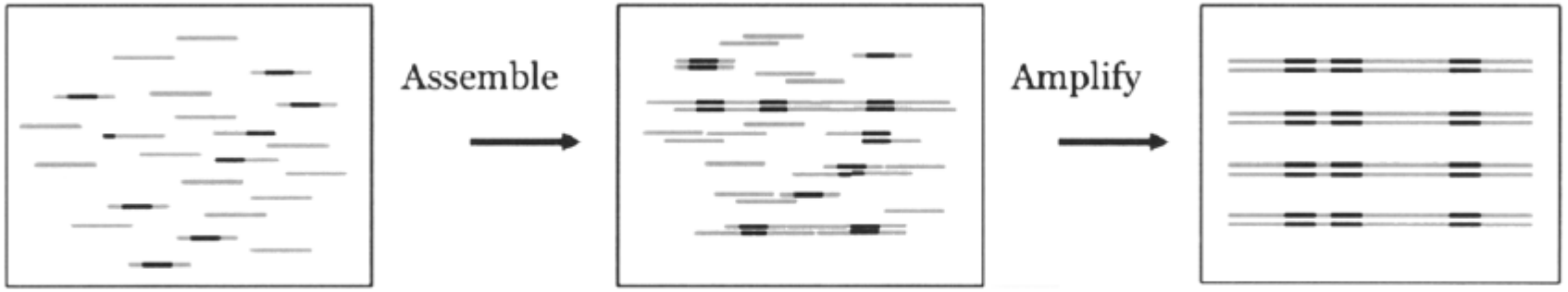
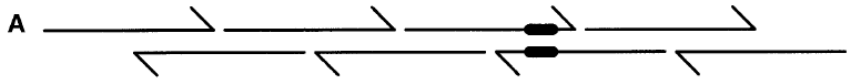
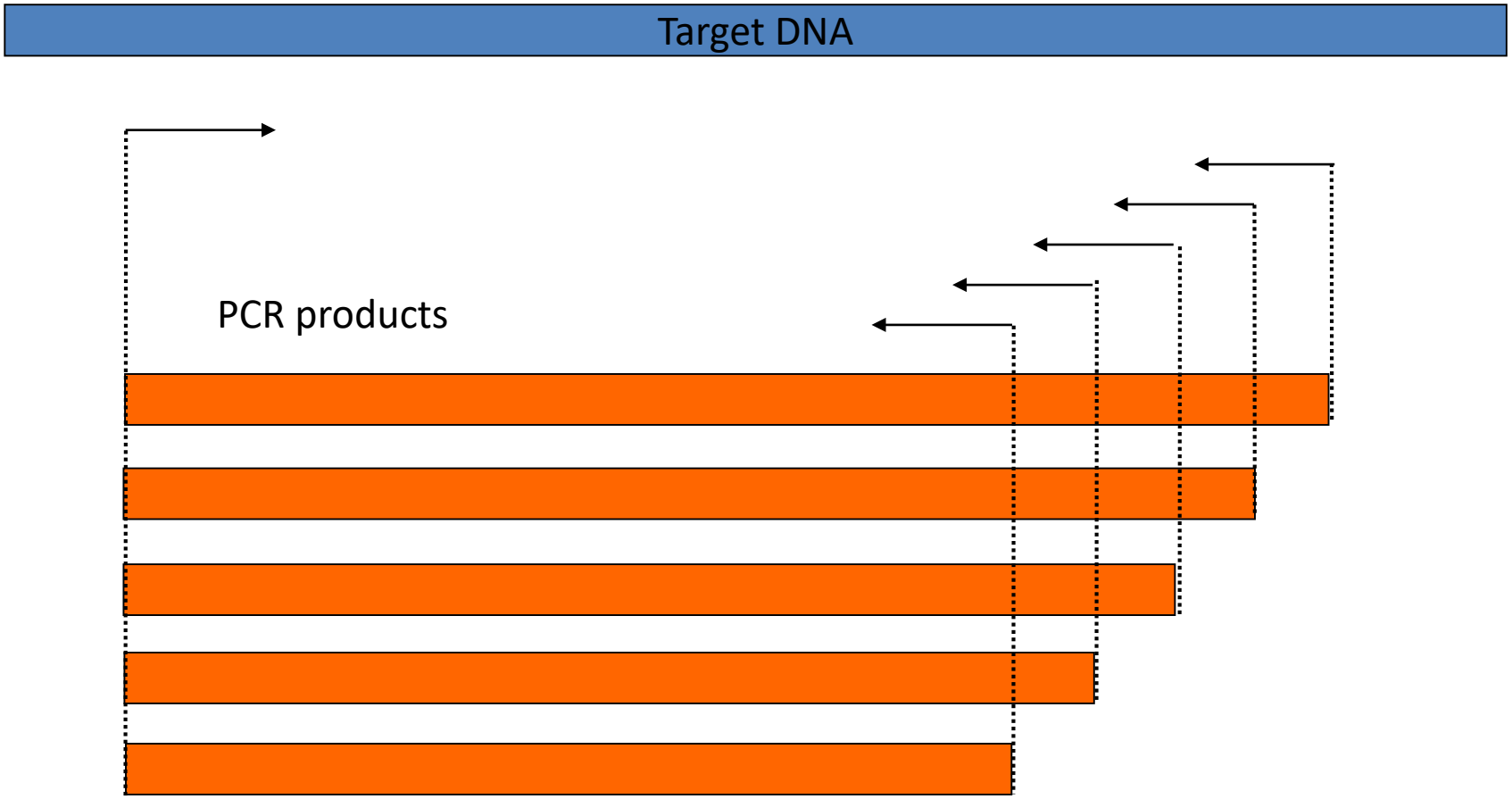


Fig. 1. Schematic of gene assembly mutagenesis. A population of gene variants is assembled from short oligonucleotides encoding both strands of the gene and containing degenerate bases at the targeted positions (shown in bold). Following assembly PCR, the full length gene variants are amplified using the outside primers.



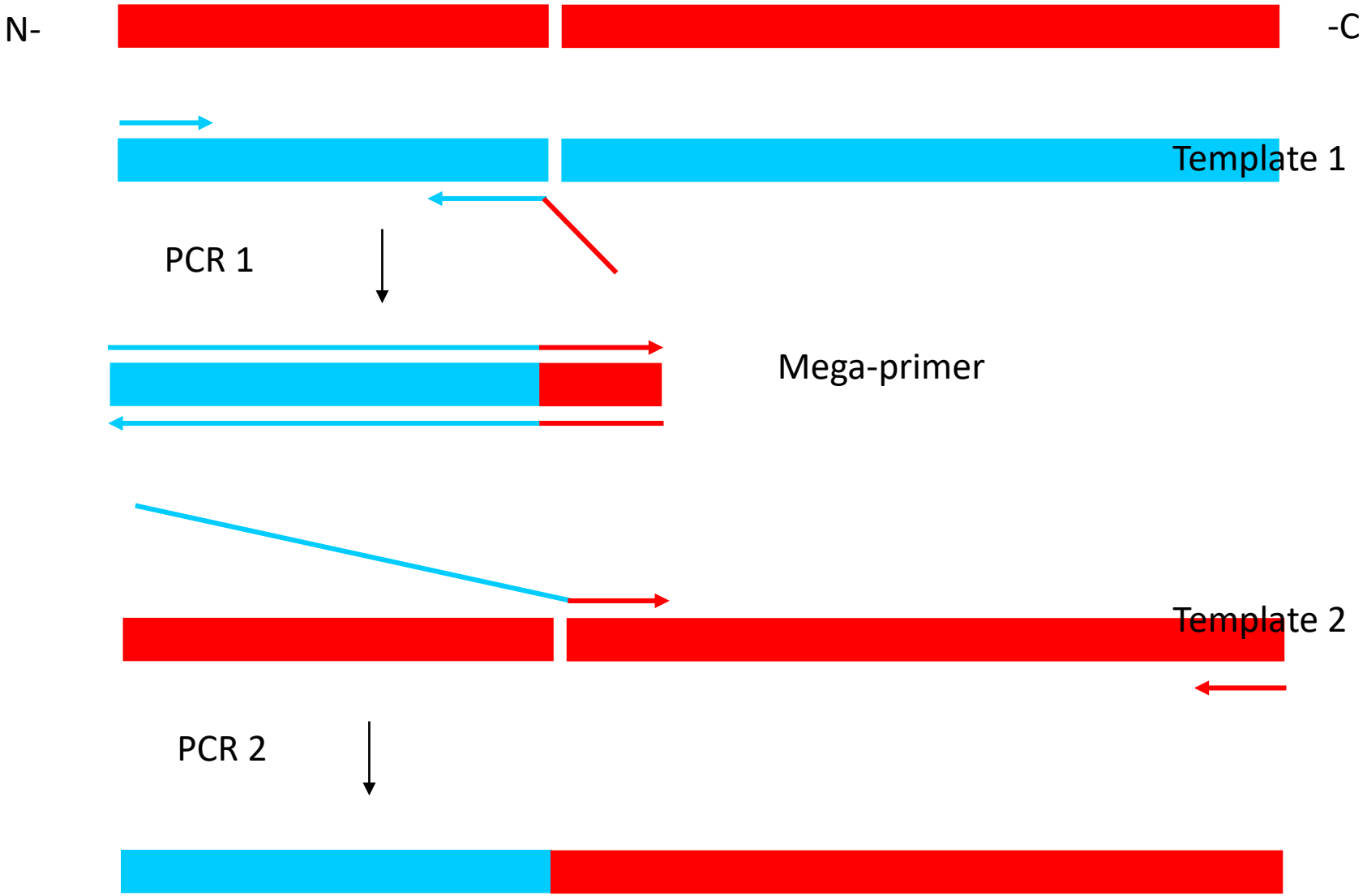
7–9 point and 0.4–1.3 frame-shifted mutations per kilobase of DNA

PCR-mediated deletion mutagenesis



Oligonucleotide design allows precision in deletion positions

Domain swapping using “megaprimers” (overlapping PCR)

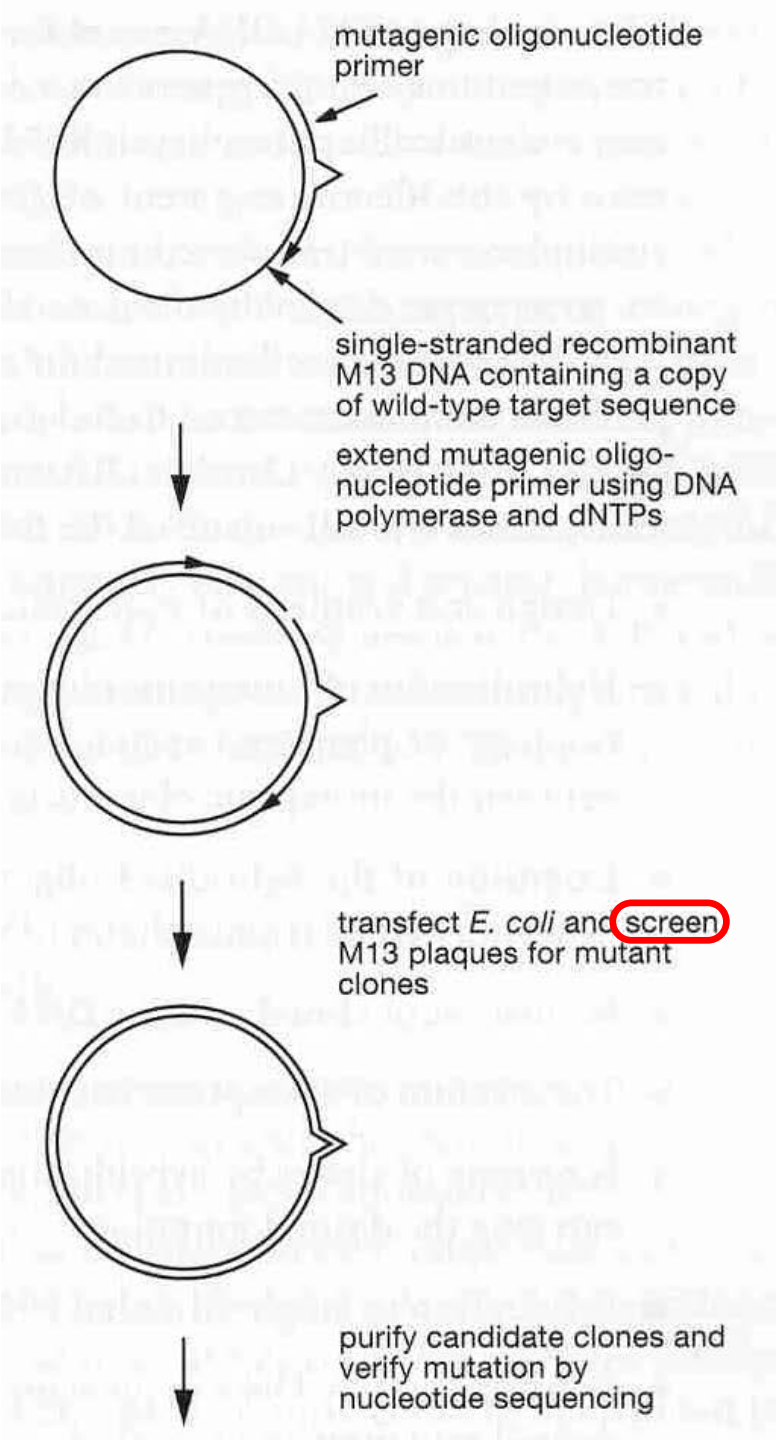


Site-directed mutagenesis: primer extension method

Drawbacks:

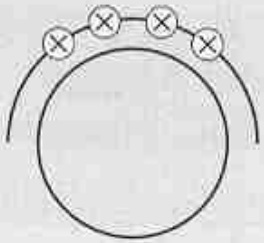
-- both mutant and wild type versions of the gene are made following transfection--lots of screening required, or tricks required to prevent replication of wild type strand

-- requires single-stranded, circular template DNA



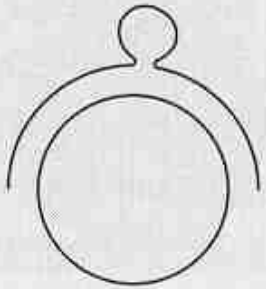
Alternative primer extension mutagenesis techniques

Multiple point mutations



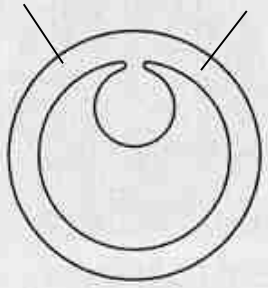
Mutant oligonucleotide with multiple (four) single base pair mismatches

Insertion mutagenesis



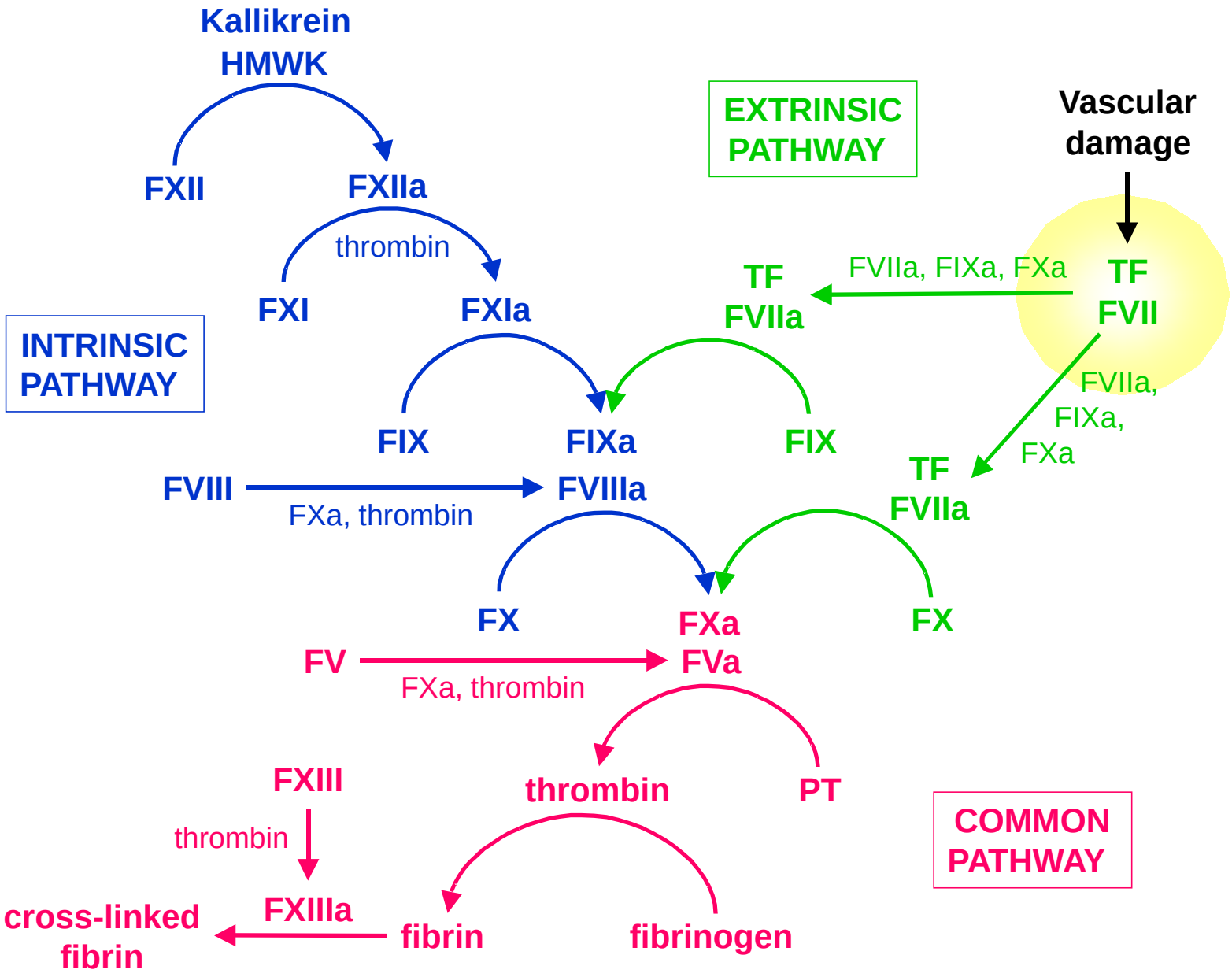
Mutant oligonucleotide carrying a sequence to be inserted sandwiched between two regions with sequences complementary to sites on either sides of the target site in the template

Deletion mutagenesis

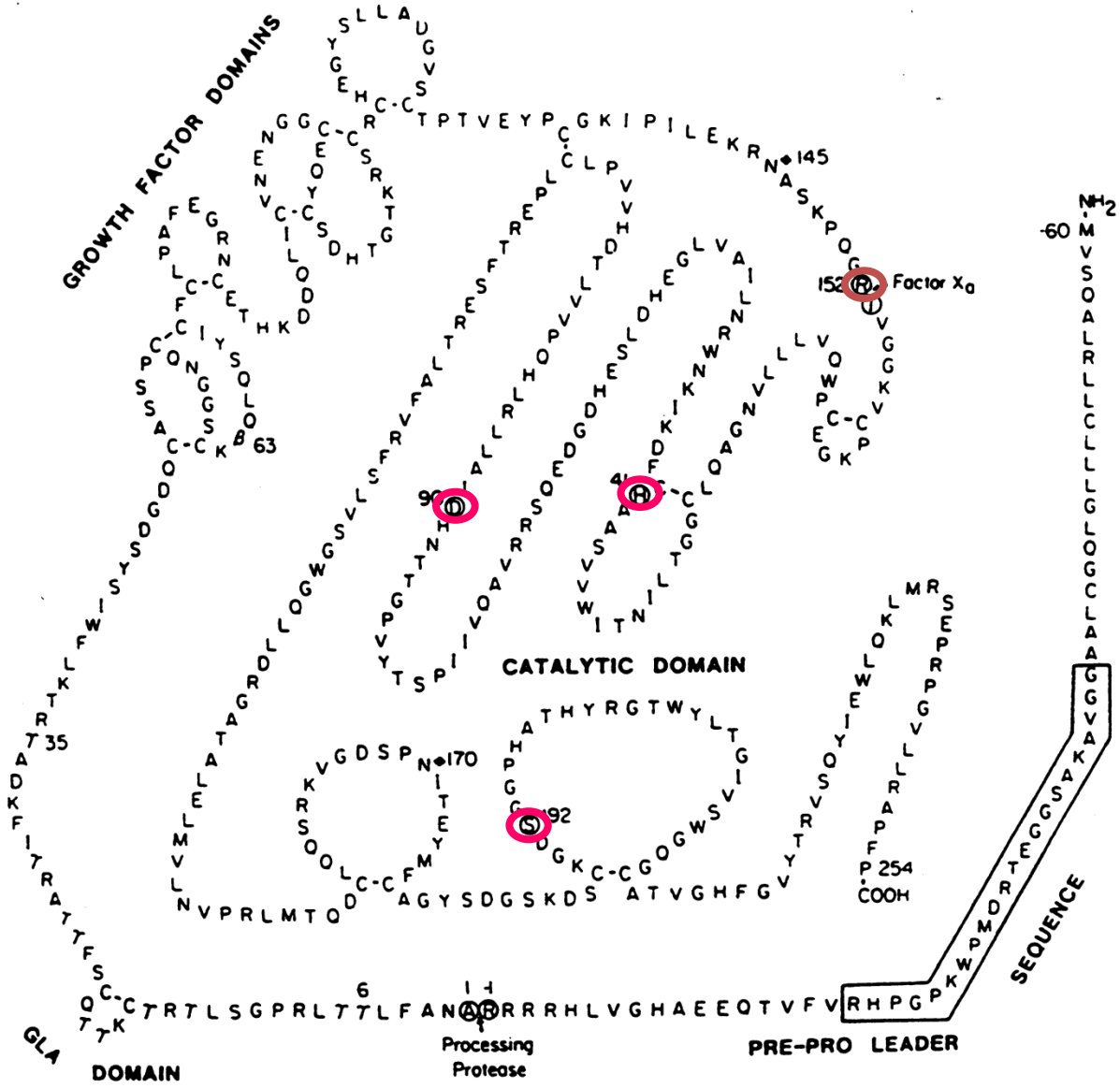


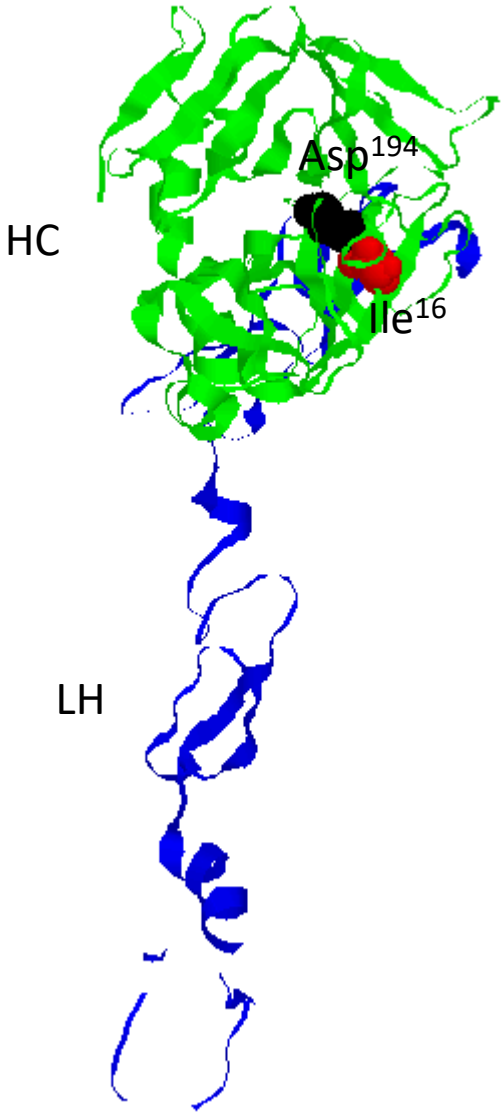
Mutant oligonucleotide spanning the region to be deleted, binding to two separate sites, one on either side of the target

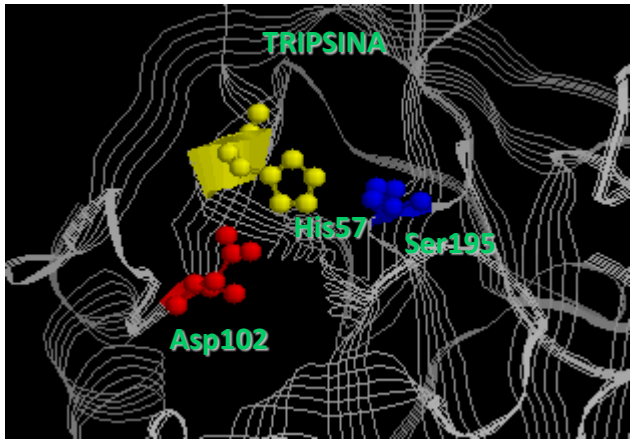
(1)
Rational design of coagulation factor VIIa variants with substantially increased intrinsic activity.

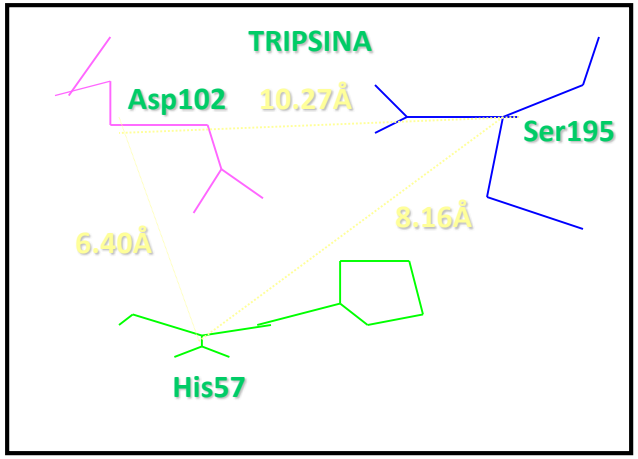

Kallikrein
HMWK
FXII
FXIIa
thrombin
FXI
FXIa
INTRINSIC PATHWAY
EXTRINSIC PATHWAY
Vascular damage
TF
FVIIa
FVIIa, FIXa, FXa
TF
FVII
FVIIa,
FIXa,
FXa
FIX
FIXa
FIX
FVIII
FVIIIa
FXa, thrombin
TF
FVIIa
FX
FXa
FX
FV
FVa
FXa, thrombin
FXIII
thrombin
PT
COMMON PATHWAY
thrombin
FXIIIa
cross-linked fibrin
fibrin
fibrinogen

FVII


GROWTH FACTOR DOMAINS
CATALYTIC DOMAIN
Factor Xa
SEQUENCE
GLA
DOMAIN
Processing
Protease
PRE-PRO LEADER
COOH


Asp194
HC
Ile16
LH
FVIIa


TRIPSINA
His57
Ser195
Asp102


TRIPSINA
Asp102
10.27Å
Ser195
8.16Å
6.40Å
His57

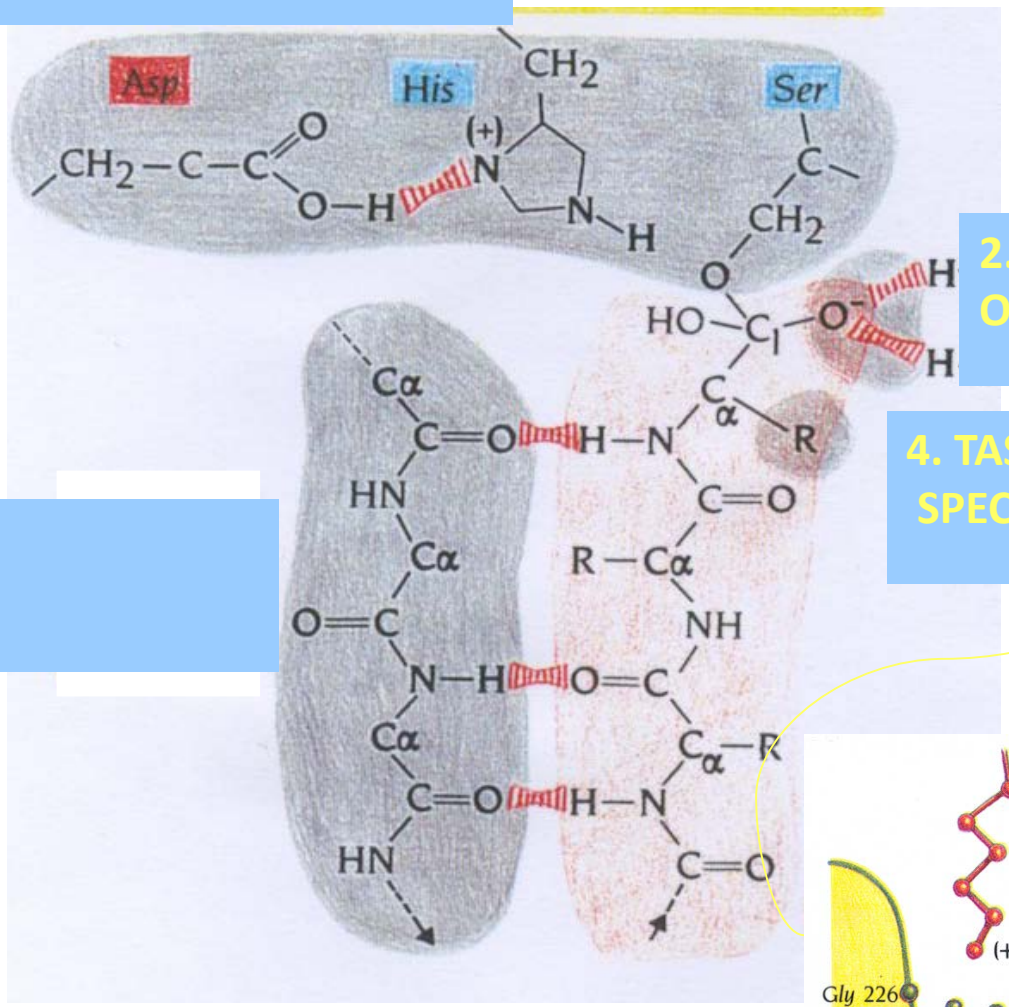
CARATTERISTICHE DEL DOMINIO SERIN PROTEASICO

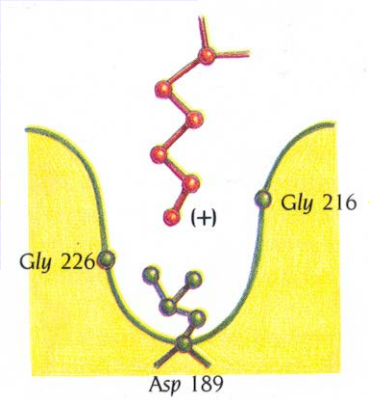
1. TRIADE CATALITICA

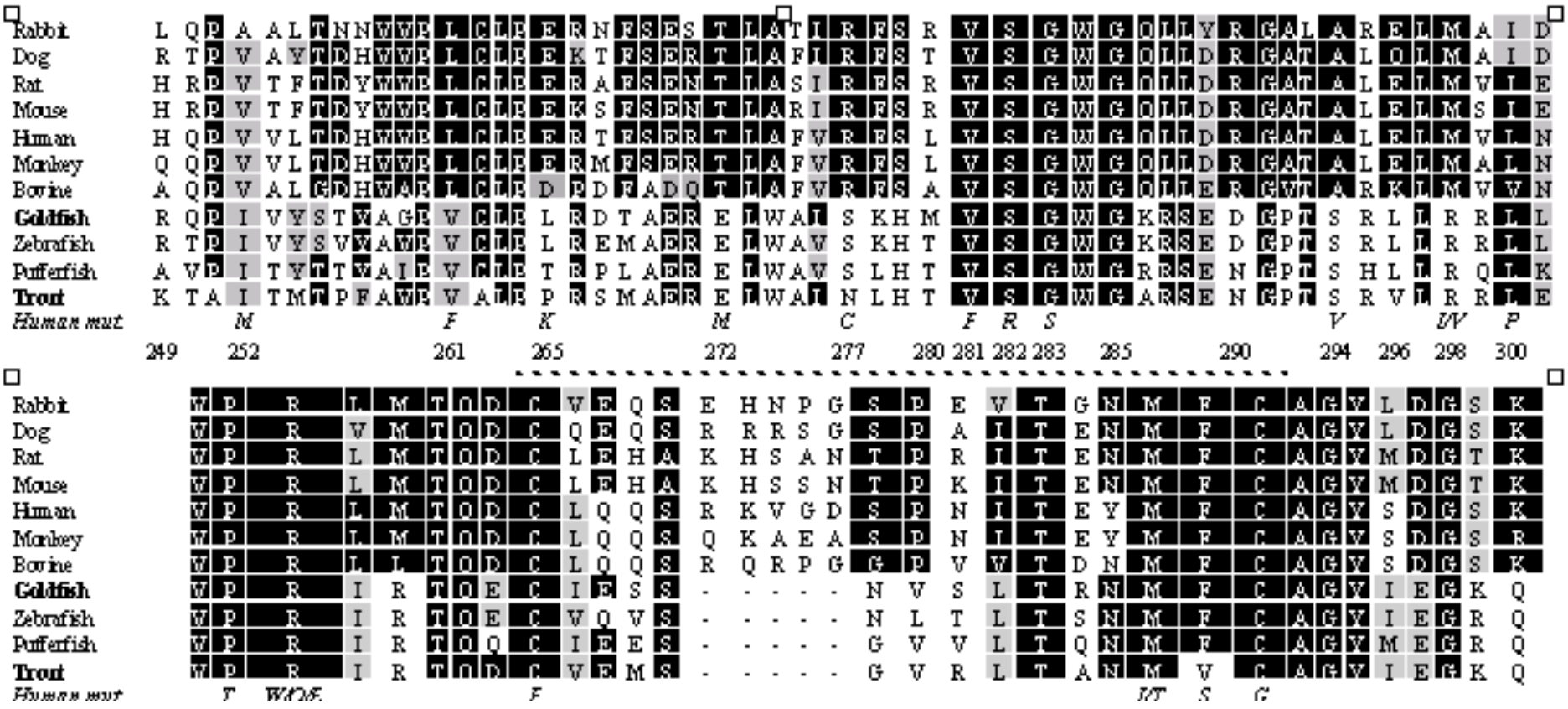
2. INCAVO OSSIANIONICO

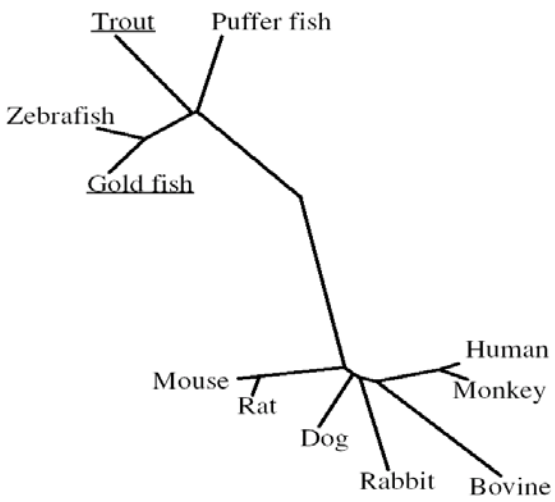
3. SITO DI LEGAME ASPECIFICO

4. TASCA DI SPECIFICITA'

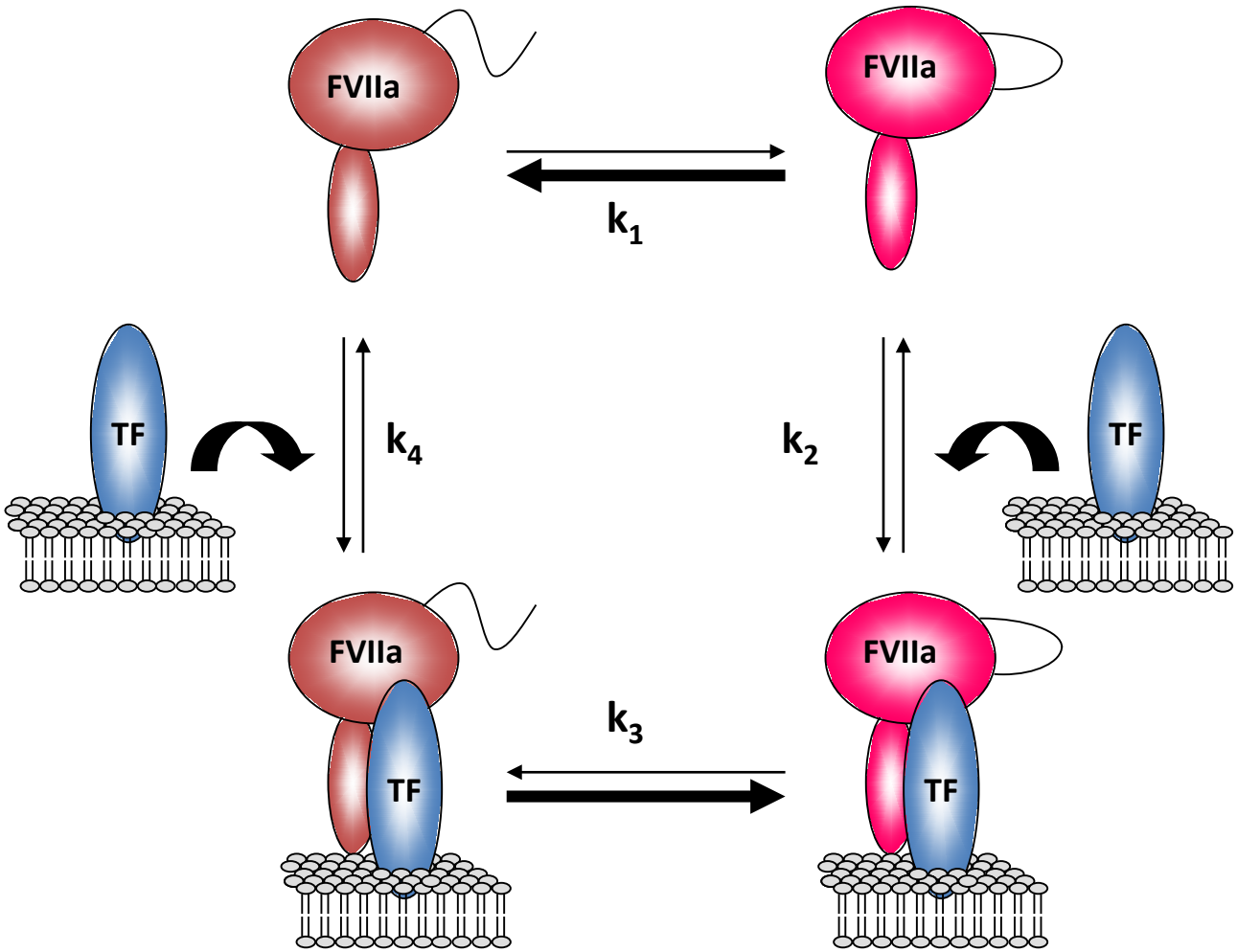




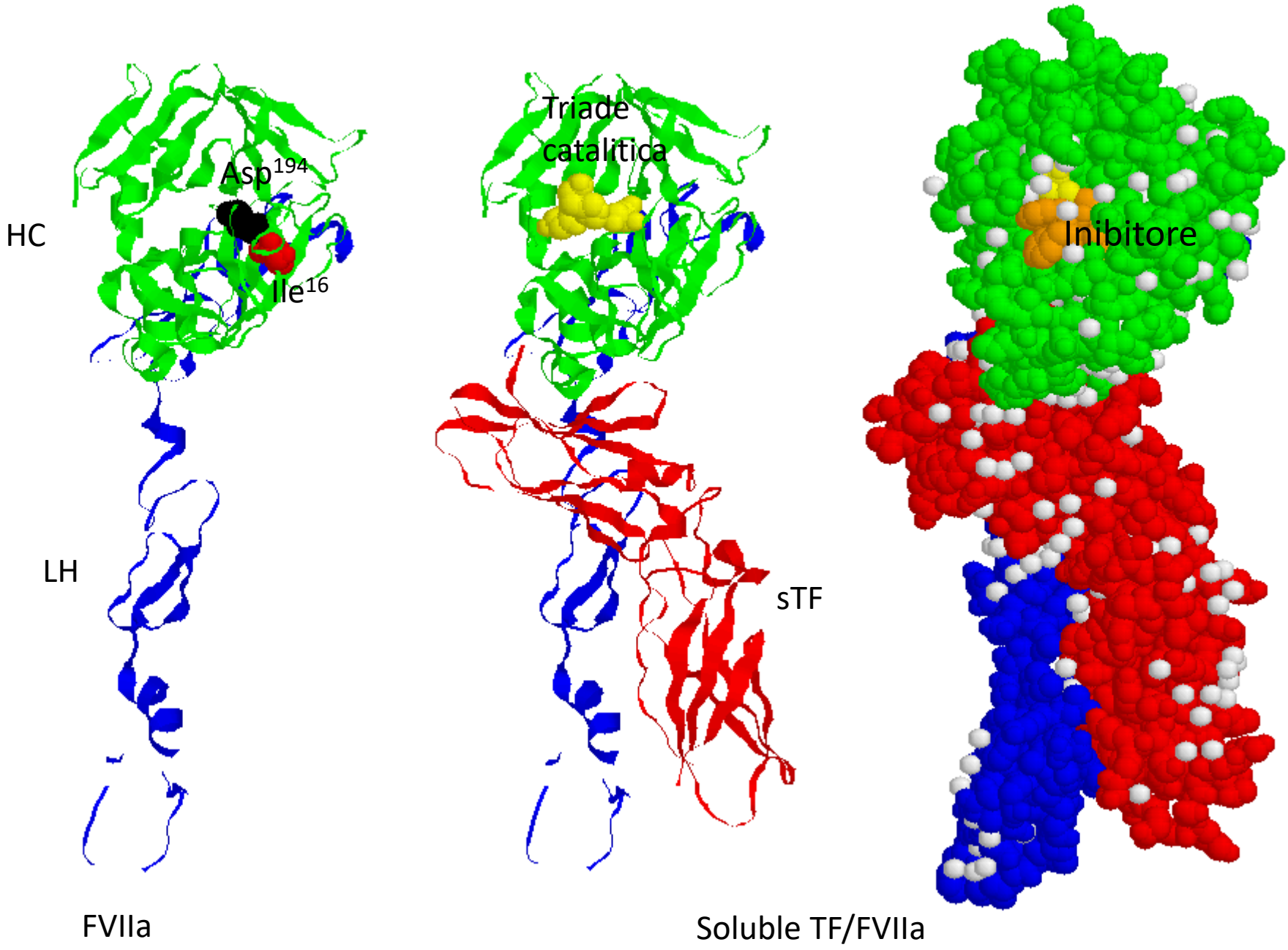


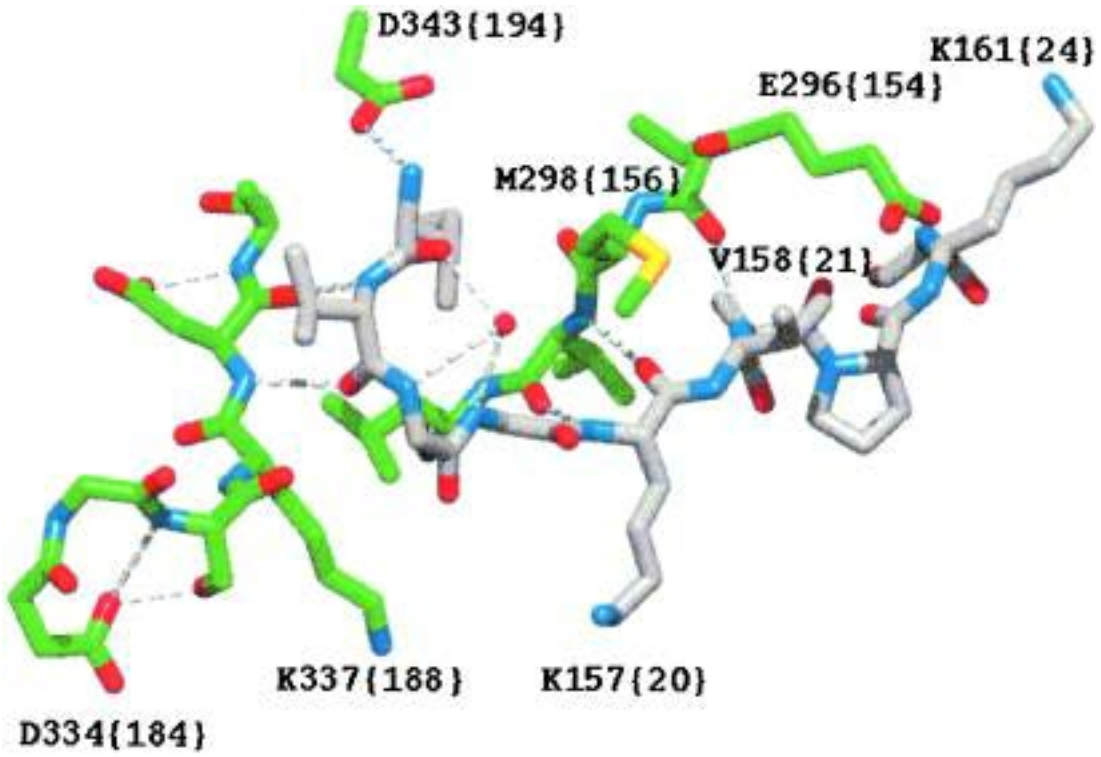


	Trout	Gold fish	Zebrafish	Puffer fish	Rat	Human	Mouse	Rabbit	Monkey	Dog	Bovine
Trout		***71***	***73***	***72***	***48***	***47***	***48***	***48***	***47***	***48***	***48***
Gold fish			***84***	***74***	***47***	***49***	***47***	***48***	***49***	***47***	***47***
Zebrafish				*74*	*45*	*51*	*45*	*47*	*51*	*48*	*47*
Puffer fish					*47*	*49*	*47*	*45*	*48*	*48*	*47*
Rat						74	93	71	71	71	60
Human							73	71	92	76	70
Mouse								71	71	74	60
Rabbit									71	74	66
Monkey										74	67
Dog											66

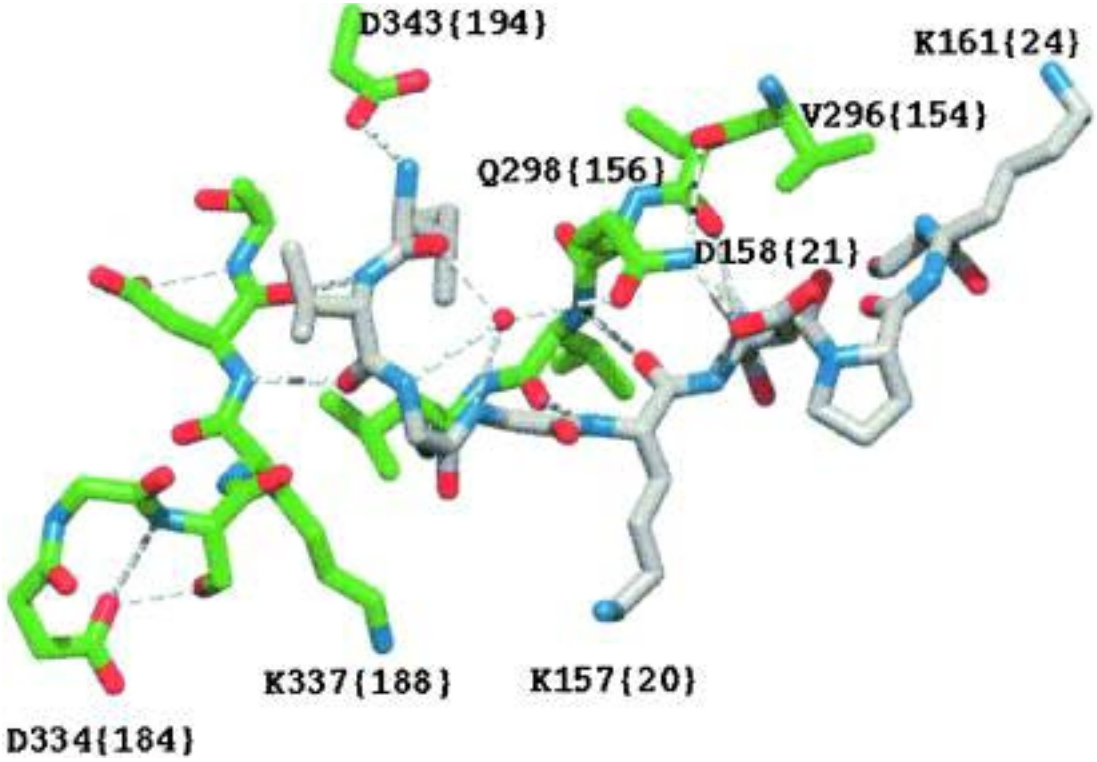


Il complesso Xasico


HC
Asp194
Ile16
LH
FVIIa
Triade
catalitica
sTF
Inibitore
Soluble TF/FVIIa



Activation pocket region of FVIIa. The structure is from the complex between FVIIa and TF. The carbon atoms of N-terminal Ile-153 {16} to Lys-161 are shown in gray and those of the amino acids constituting part of the activation pocket are in green. The water molecule (shown as a red sphere) interacting with main chain atoms of Gly-155 {18} and Gly-156 {19} lacks hydrogen bonds to the side chain of Met-298 {156}.



Activation pocket region of FVIIa after mutating the residues in positions 158 {21}, 296 {154}, and 298 {156} to those occupying the corresponding positions in thrombin (Asp, Val and Gln, respectively). The backbone structure (3) and coloring scheme are the same as in Fig. 1. The introduced side chains are oriented as in the thrombin structure. Note that a hydrogen bond network between the water molecule, Gln-298 {156} and Asp-158 {21} is established.

Table 2. Kinetics of the interaction between the FVIIa variants and sTF

FVIIa variant	k_{on}, $\times 10^5$ $M^{-1}\cdot s^{-1}$	k_{off}, $\times 10^{-3}$ s^{-1}	K_d, nM
Wild-type FVIIa	9.3 ± 0.3	2.1 ± 0.2	2.3
K337A-FVIIa	8.9 ± 1.1	1.7 ± 0.2	1.9
$FVIIa_{IIa}$	9.6 ± 1.3	1.4 ± 0.2	1.5

Values are means ± SD ($n = 3$). K_d (k_{off}/k_{on}) values are calculated from the means.

Proc. Natl. Acad. Sci. USA. 2001 November; 98 (24): 13583–13588

In the presence of Tissue factor the activity of variants was comparable or slightly increased as compared to wtFVIIa

Table 1. Amidolytic activity of FVIIa variants and kinetic parameters of FX activation

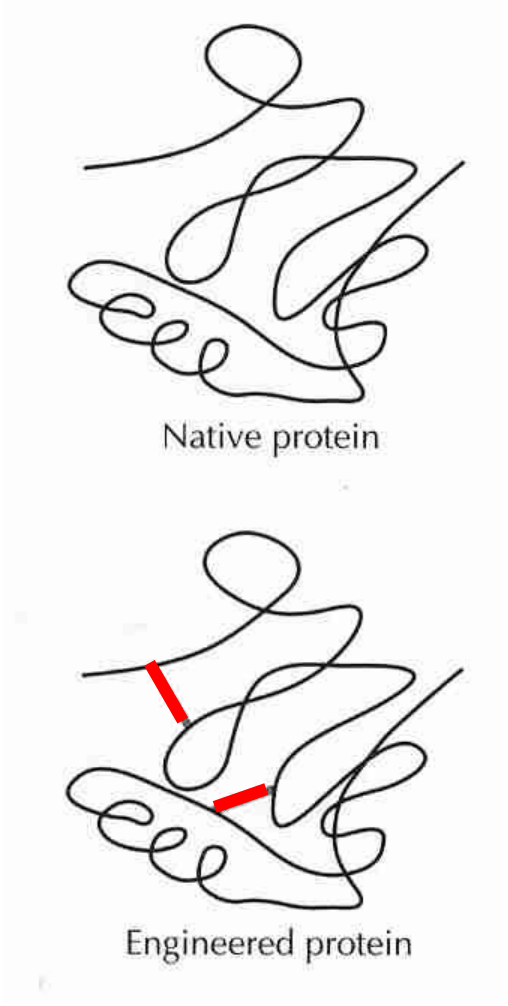
FVIIa variant	Amidolytic activity	FX activation		
	Mutant/wt	K_m, μM	k_{cat}, × 10^{-3} s^{-1}	k_{cat}/K_m, $M^{-1}\cdot s^{-1}$
wt-FVIIa		2.9 ± 0.5	0.094 ± 0.010	32
K337A-FVIIa	3.9 ± 0.2	2.0 ± 0.6	0.28 ± 0.05	140
L305V-FVIIa	3.2 ± 0.2	2.9 ± 0.5	0.48 ± 0.06	170
L305V/K337A-FVIIa	7.2 ± 0.8	1.9 ± 0.4	0.37 ± 0.04	200
$FVIIa_{IIa}$	7.8 ± 0.3	2.3 ± 0.7	2.6 ± 0.5	1,200
K337A-$FVIIa_{IIa}$	11.0 ± 0.2	2.4 ± 0.1	4.4 ± 0.2	1,800
L305V-$FVIIa_{IIa}$	6.7 ± 0.2	2.7 ± 0.1	4.2 ± 0.2	1,600
L305V/K337A-$FVIIa_{IIa}$	11.5 ± 0.3	2.1 ± 0.2	6.8 ± 0.2	3,200
M298Q-FVIIa	3.4 ± 0.2	2.4 ± 0.2	0.52 ± 0.07	220

All values are means ± SD. The amidolytic activity is given as the ratio between the activity of the mutant and the activity of wild-type (wt) FVIIa in the presence of 1 mM S-2288 ($n = 3$). The k_{cat}/K_m values are calculated from the means ($n = 2$).

Directed mutagenesis

- Make changes in amino acid sequence based on rational decisions
- Structure known? Mutate amino acids in any part of protein thought to influence activity/stability/solubility etc.
- Protein with multiple family members? Mutate desired protein in positions that bring it closer to another family member with desired properties

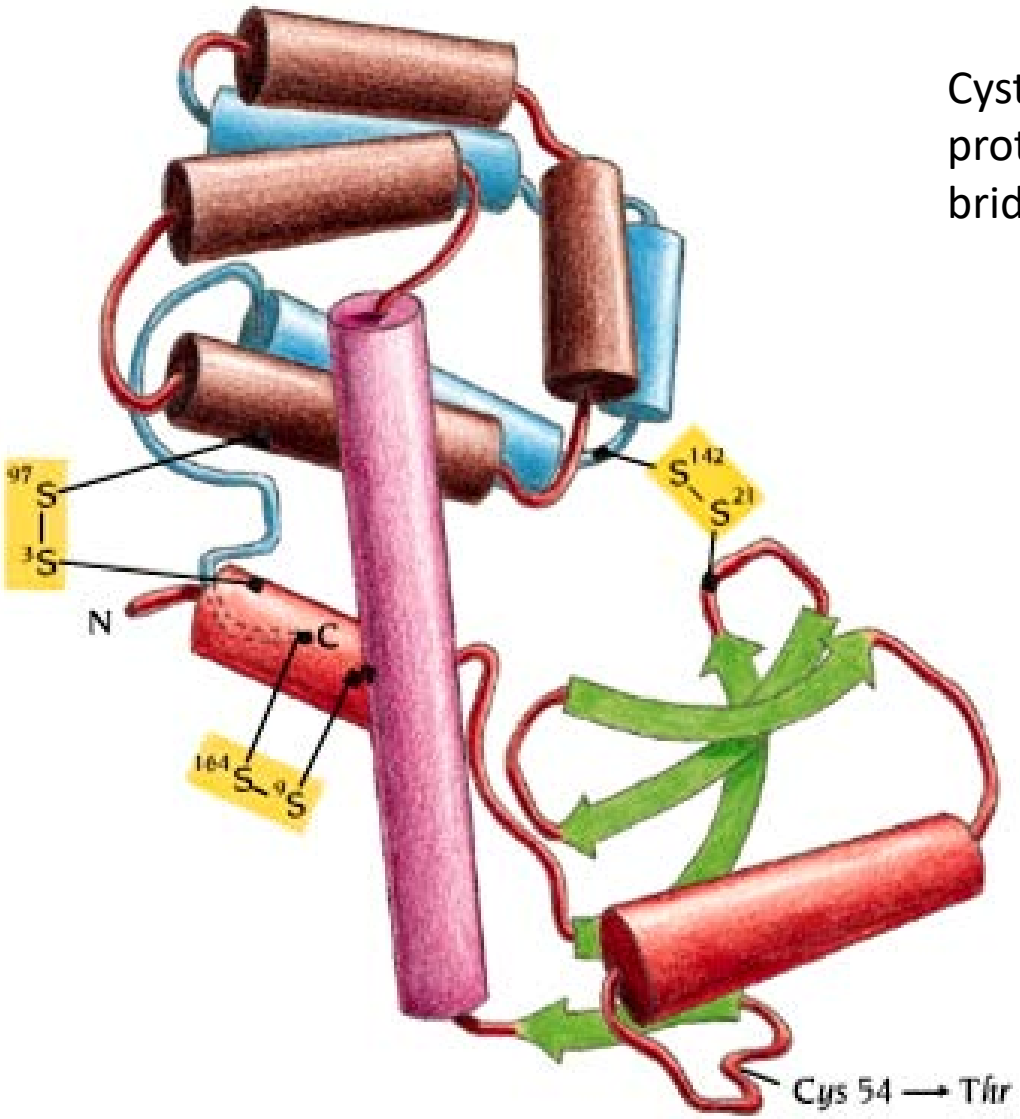
An example of directed mutagenesis



T4 lysozyme: structure known

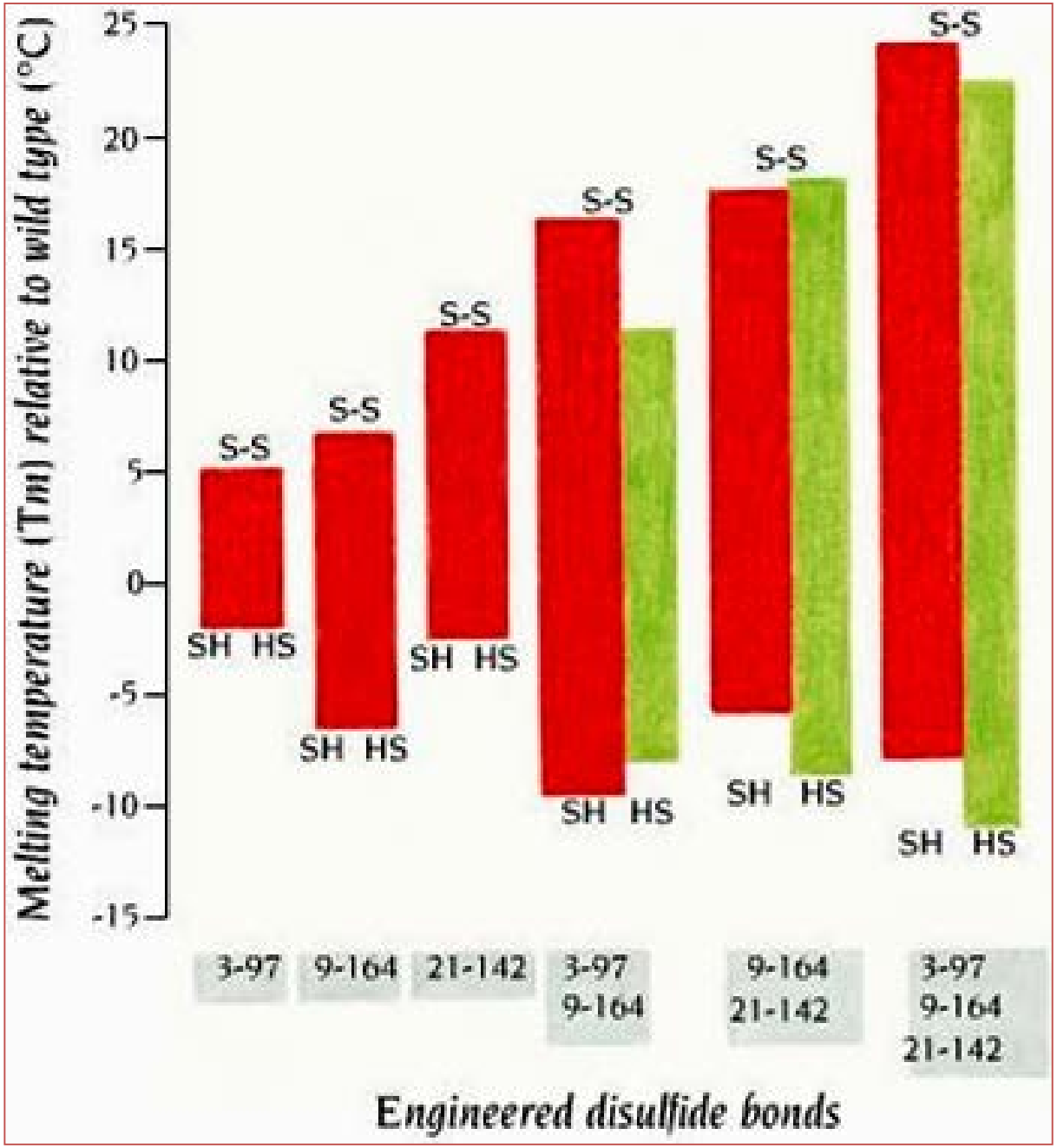
Can it be made more stable by the addition of pairs of cysteine residues (allowing disulfide bridges to form?) without altering activity of the protein?

T4 lysozyme: a model for stability studies



Cysteines were added to areas of the protein in close proximity--disulfide bridges could form

More disulfides, greater stabilization at high T



Bottom of bar: melting temperature under reducing condtions

Top of bar:
Melting temperature under oxidizing conditions

Green bars: if the effects of individual S-S bonds were added together

Table 8.2 Properties of T4 lysozyme and six engineered variants

Enzyme	Amino acid at position:							No. of -S-S-	% Activity	T_m (°C)
	3	9	21	54	97	142	164			
wt	Ile	Ile	Thr	Cys	Cys	Thr	Leu	0	100	41.9
pwt	Ile	Ile	Thr	Thr	Ala	Thr	Leu	0	100	41.9
A	Cys	Ile	Thr	Thr	Cys	Thr	Leu	1	96	46.7
B	Ile	Cys	Thr	Thr	Ala	Thr	Cys	1	106	48.3
C	Ile	Ile	Cys	Thr	Ala	Cys	Leu	1	0	52.9
D	Cys	Cys	Thr	Thr	Cys	Thr	Cys	2	95	57.6
E	Ile	Cys	Cys	Thr	Ala	Cys	Cys	2	0	58.9
F	Cys	Cys	Cys	Thr	Cys	Cys	Cys	3	0	65.5

Adapted from Matsumura et al., *Nature* **342**:291–293, 1989.

wt, wild-type T4 lysozyme; pwt, pseudo-wild-type enzyme; A through F, six engineered cysteine variants; -S-S-, disulfide bonds; T_m, "melting" temperature (a measure of thermostability).

Stability can be increased - but there can be a cost in activity

IRRATIONAL DESIGN

To attempt to mimic the natural processes by which protein variants arise and are tested for fitness within living systems

Directed Evolution – Random mutagenesis

-> based on the process of natural evolution

- **NO structural information required**

- **NO understanding of the mechanism required**

<u>General Procedure:</u>

Generation of genetic diversity
⇒ Random mutagenesis

Identification of successful variants
⇒ Screening and seletion

Directed Evolution Library

Even a large library -> (10^8 independent clones)

will not exhaustively encode all possible single point mutations.

Requirements would be:

20^N independend clones -> to have all possible variations in a library

(+ silent mutations)

N….. number of amino acids in the protein

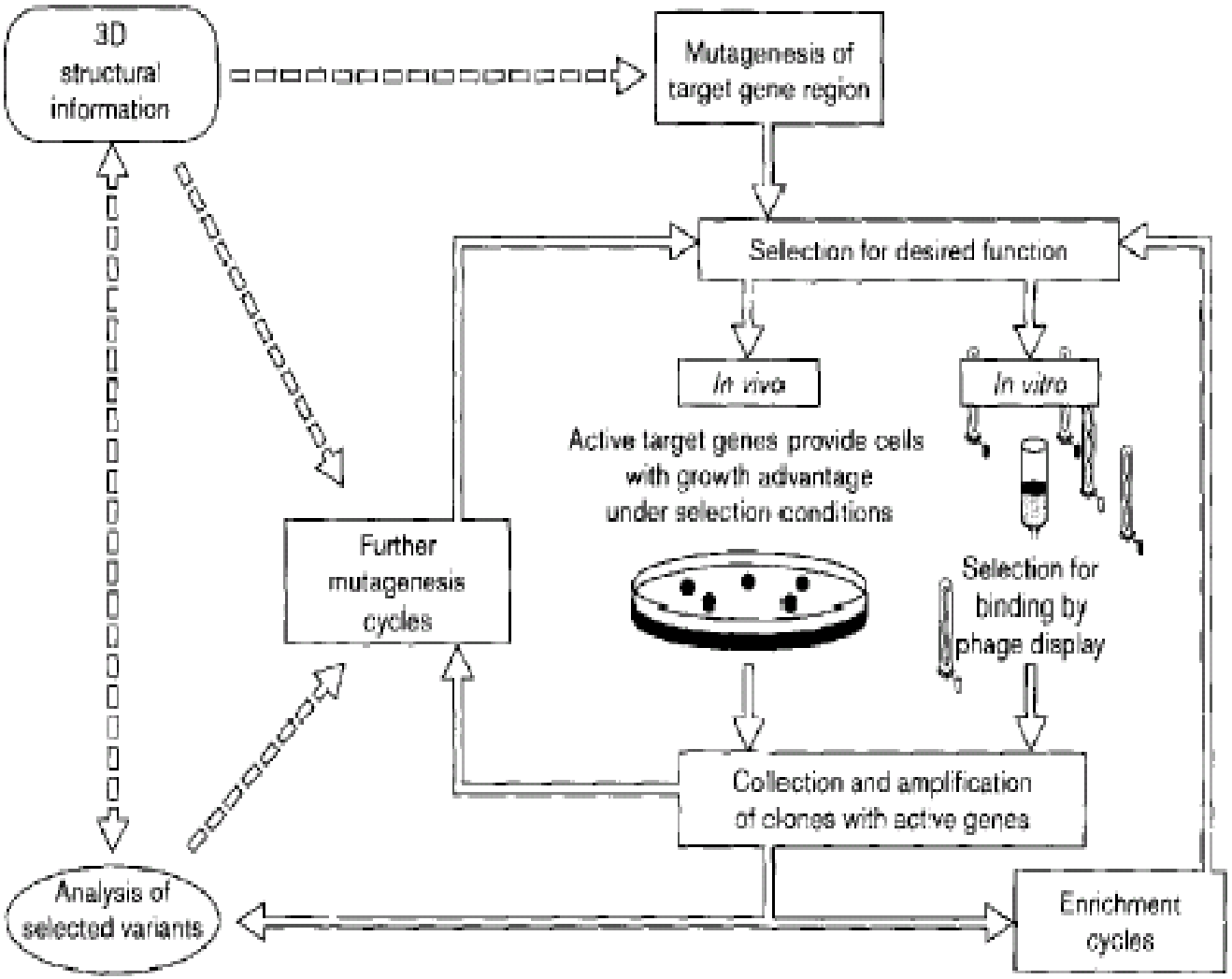
For a small protein: -> Hen egg-white Lysozyme (129 aa; 14.6 kDa)

-> library with 20^{129} ($7x\ 10^{168}$) independent clones

Consequence -> not all modifications possible

-> modifications just along an evolutionary path !!!!

The outcome of directed evolution experiments is critically dependent on how a library is screened



Selection:

only those clones that are actually desided appear

Screening:

When all members of the library are present when one chooses the best for further analysis

Limitation of Directed Evolution

1. Evolutionary path must exist - > to be successful

2. Screening method must be available

   -> You get (exactly) what you ask for!!!

   -> need to be done in -> High throughput !!!

Evolutionary Methods

- **Non-recombinative methods:**

  -> Oligonucleotide Directed Mutagenesis (saturation mutagenesis)

  -> Chemical Mutagenesis, Bacterial Mutator Strains

  -> Error-prone PCR

- **Recombinative methods** -> Mimic nature's recombination strategy

  Used for: Elimination of neutral and deleterious mutations

  -> DNA shuffling

  -> Invivo Recombination (Yeast)

  -> Random priming recombination, Staggered extention precess (StEP)

  -> ITCHY

Evolutionary Methods
Type of mutation – Fitness of mutants

Type of mutations:

⇒ Beneficial mutations (good)
⇒ Neutral mutations
⇒ Deleterious mutations (bad)

⇒ Beneficial mutations are diluted with neutral and deleterious ones

!!! Keep the number of mutations low per cycle

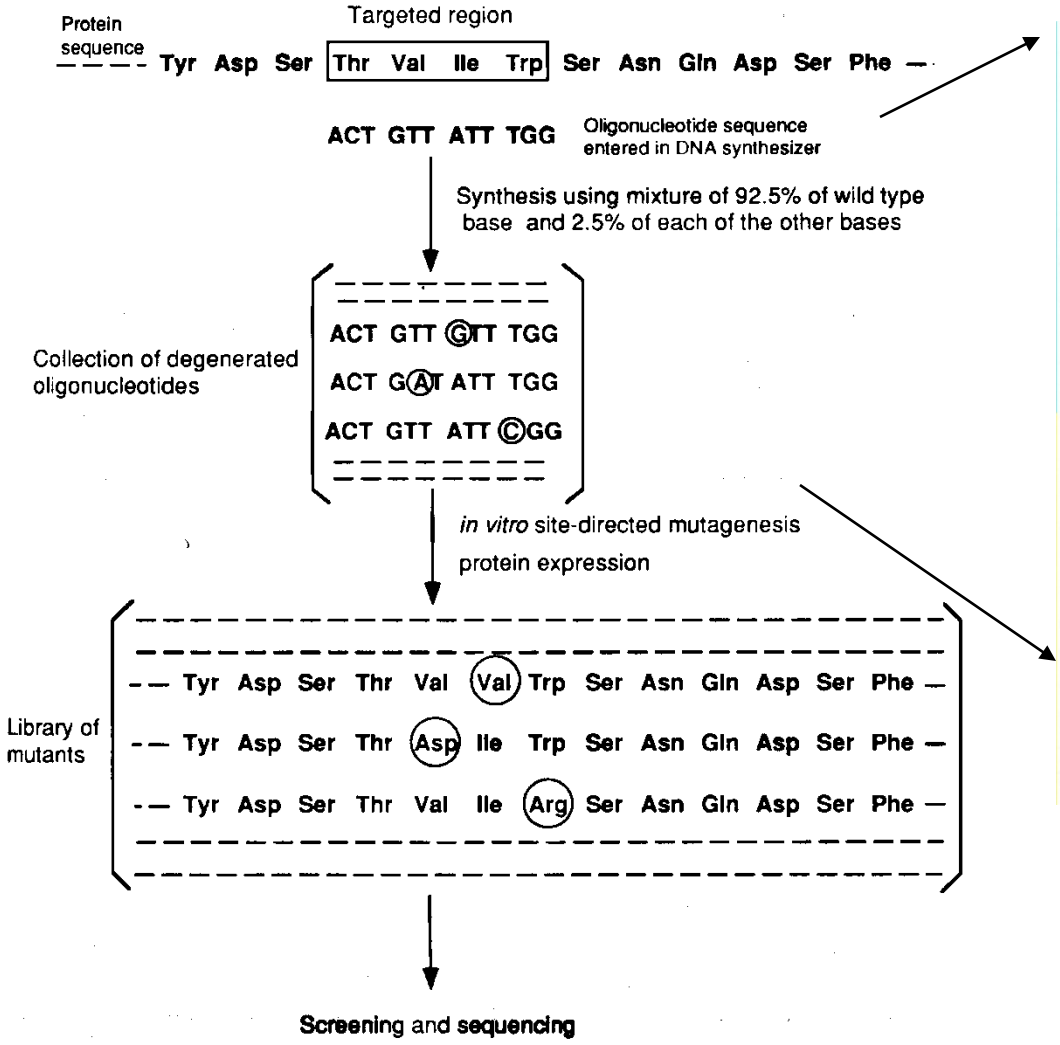
-> improve fitness of mutants!!!

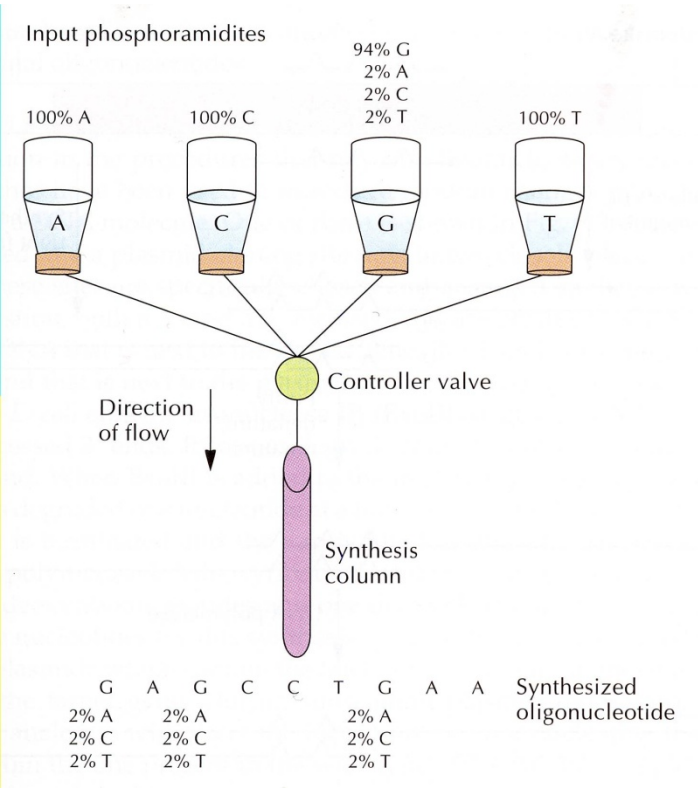
CLONAL INTEFERENCE

Competition between beneficial mutations in asexual populations
is called “Clonal Interference”
Recursive mutagenesis PCR produced essentially asexual populations within which the beneficial mutations drove each other into extintion.

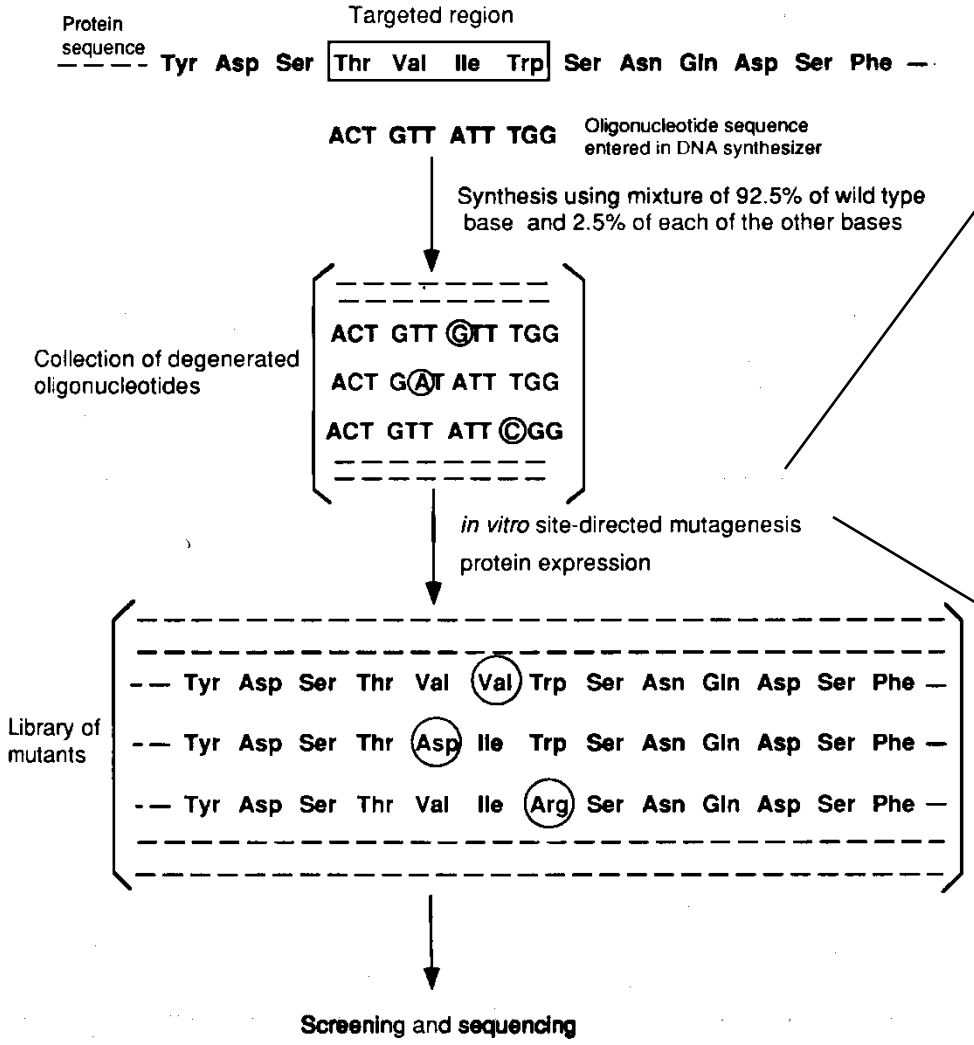
DNA shuffling (and combinatorial cassette mutagenesis) instead enable accumulation of these mutations in super-alleles

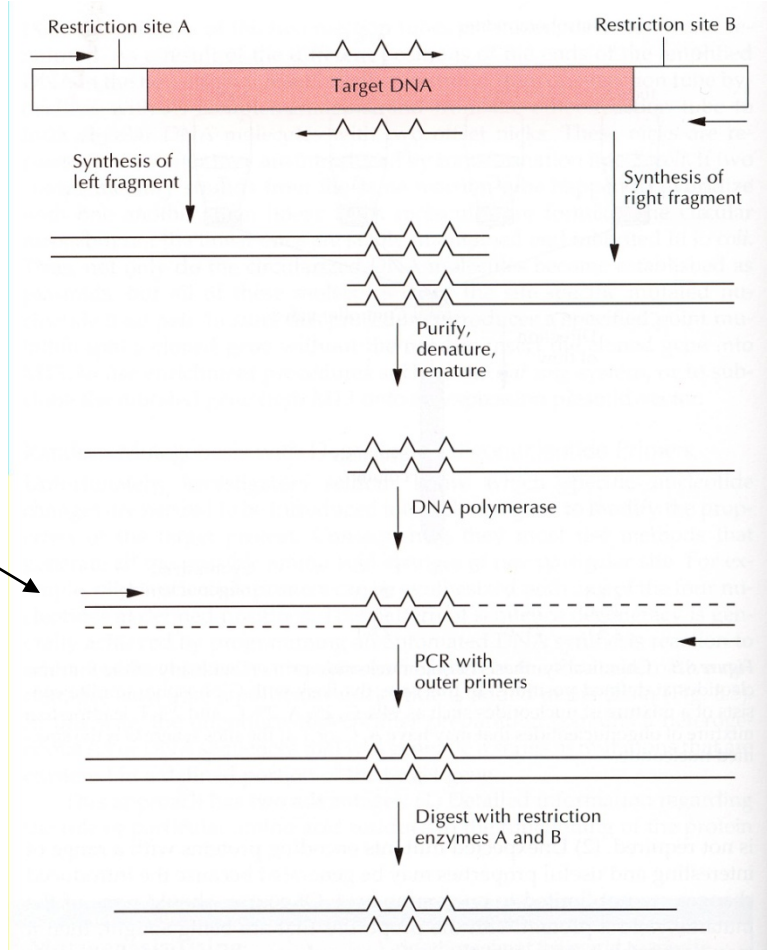
Random Mutagenesis (PCR based) with degenerated primers (saturation mutagenesis)



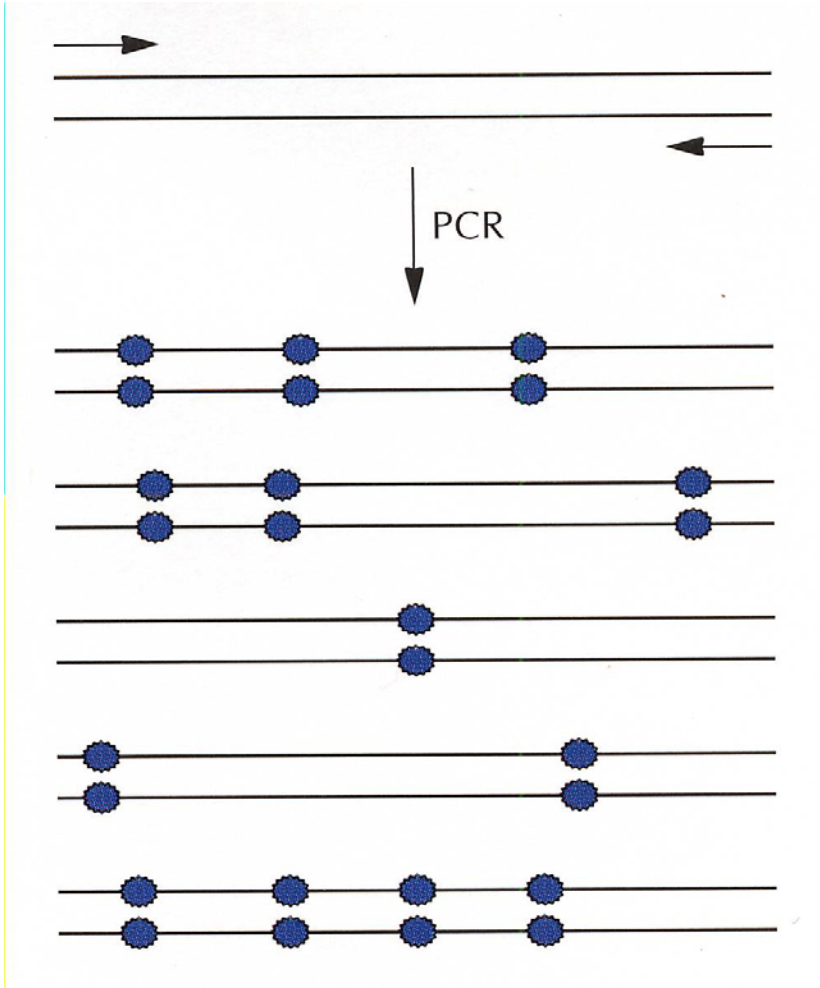


Random Mutagenesis (PCR based) with degenerated primers (saturation mutagenesis)





Random Mutagenesis (PCR based)
Error –prone PCR

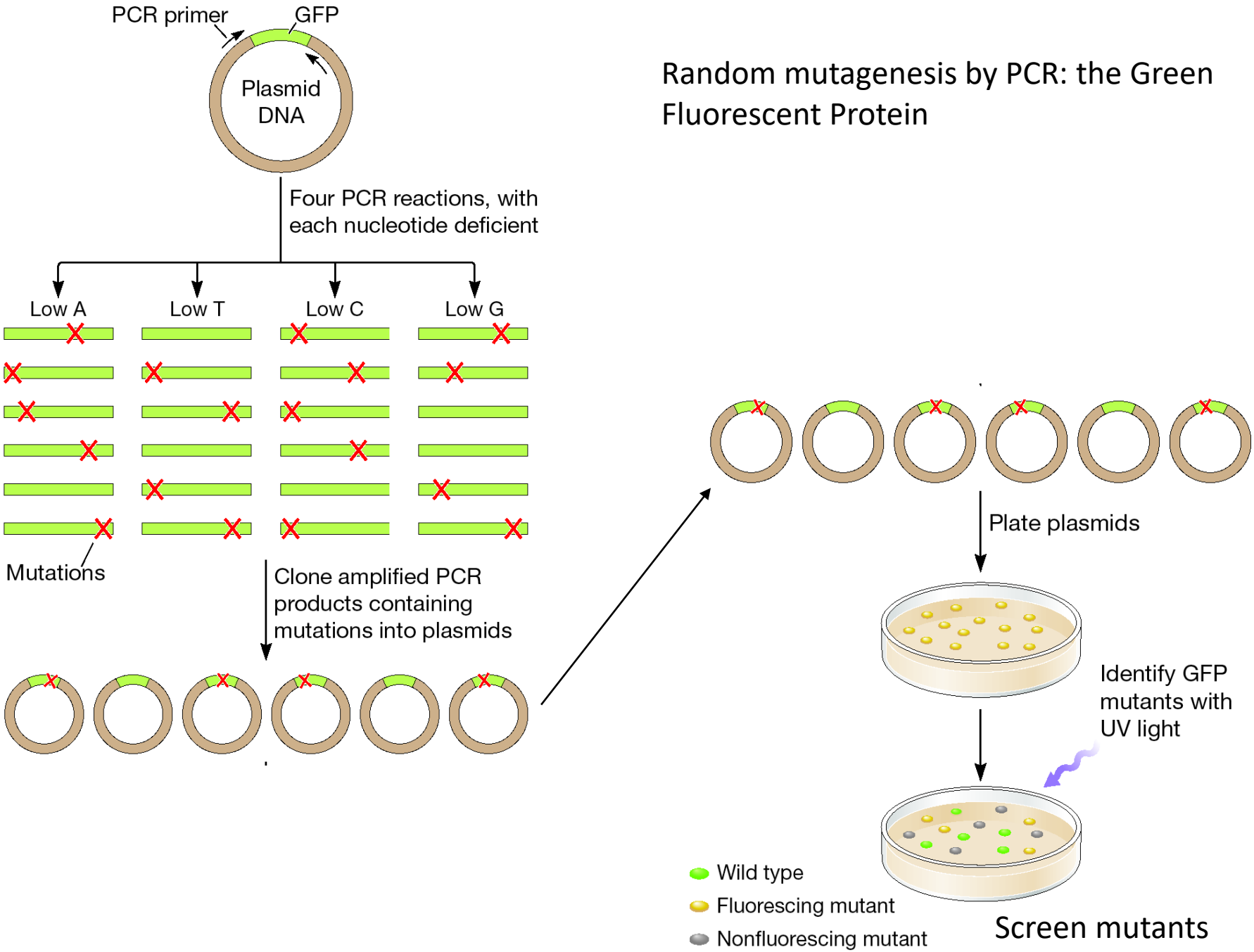


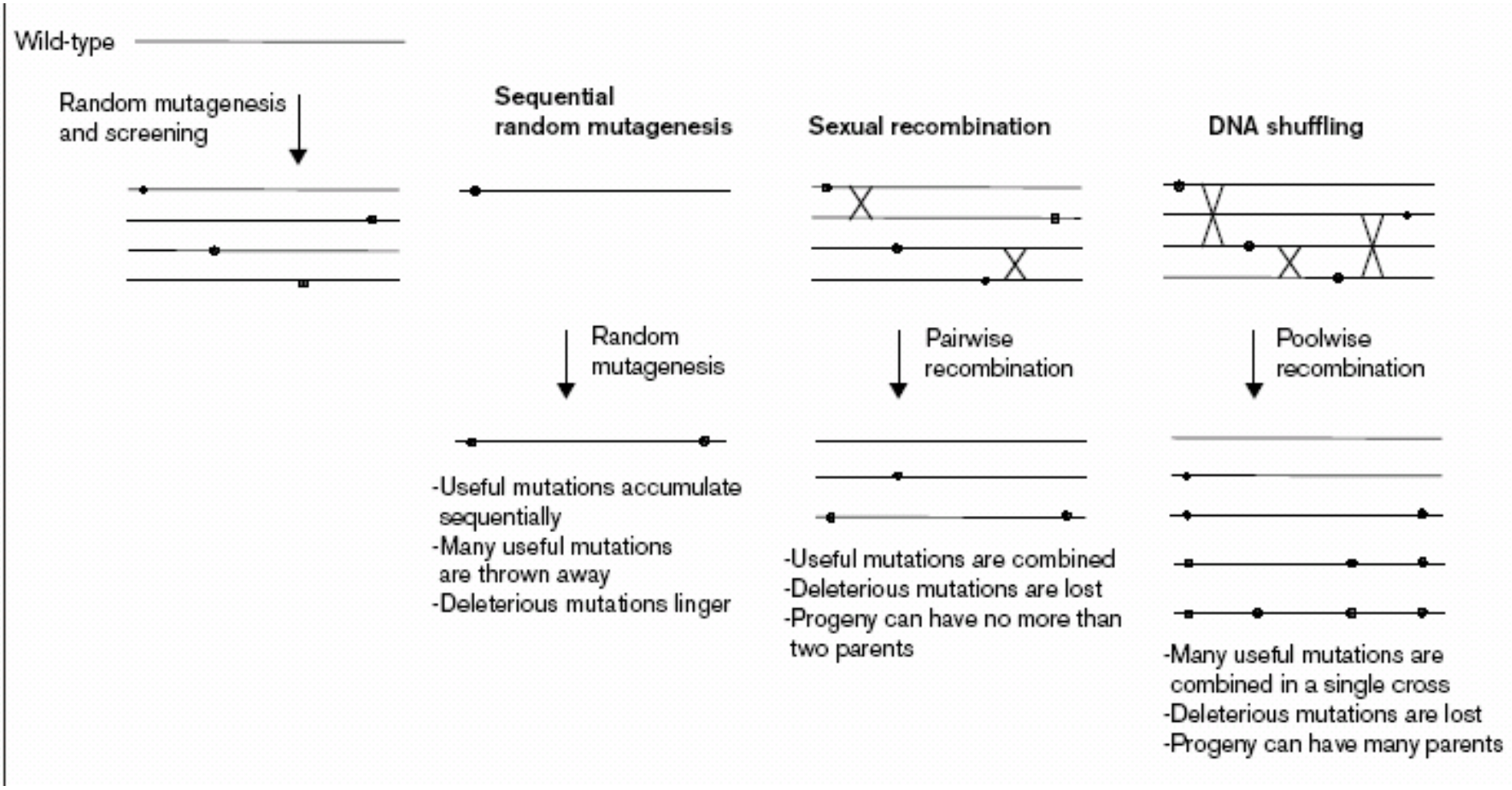
-> PCR with low fidelity !!!

Achieved by:

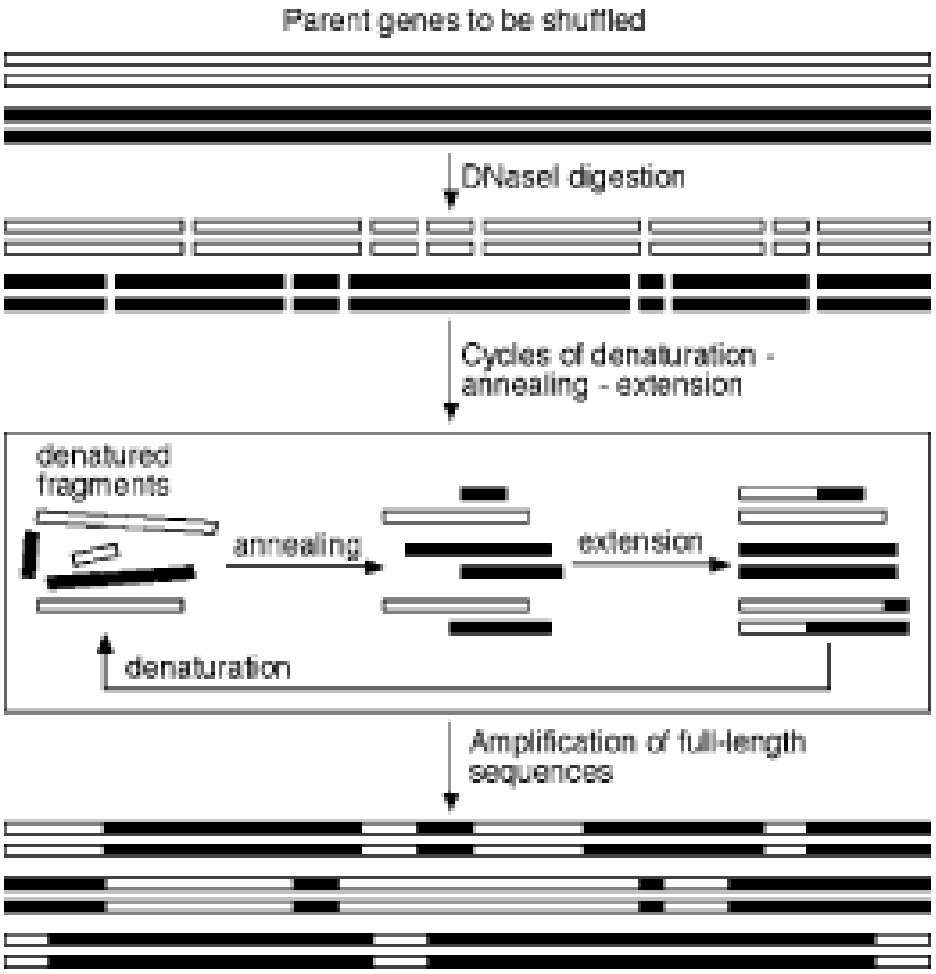
- Increased Mg2+ concentration
- Addition of Mn2+
- Not equal concentration of the four dNTPs
- Use of dITP
- Increasing amount of Taq polymerase (Polymerase with NO proof reading function)

Random mutagenesis by PCR: the Green Fluorescent Protein



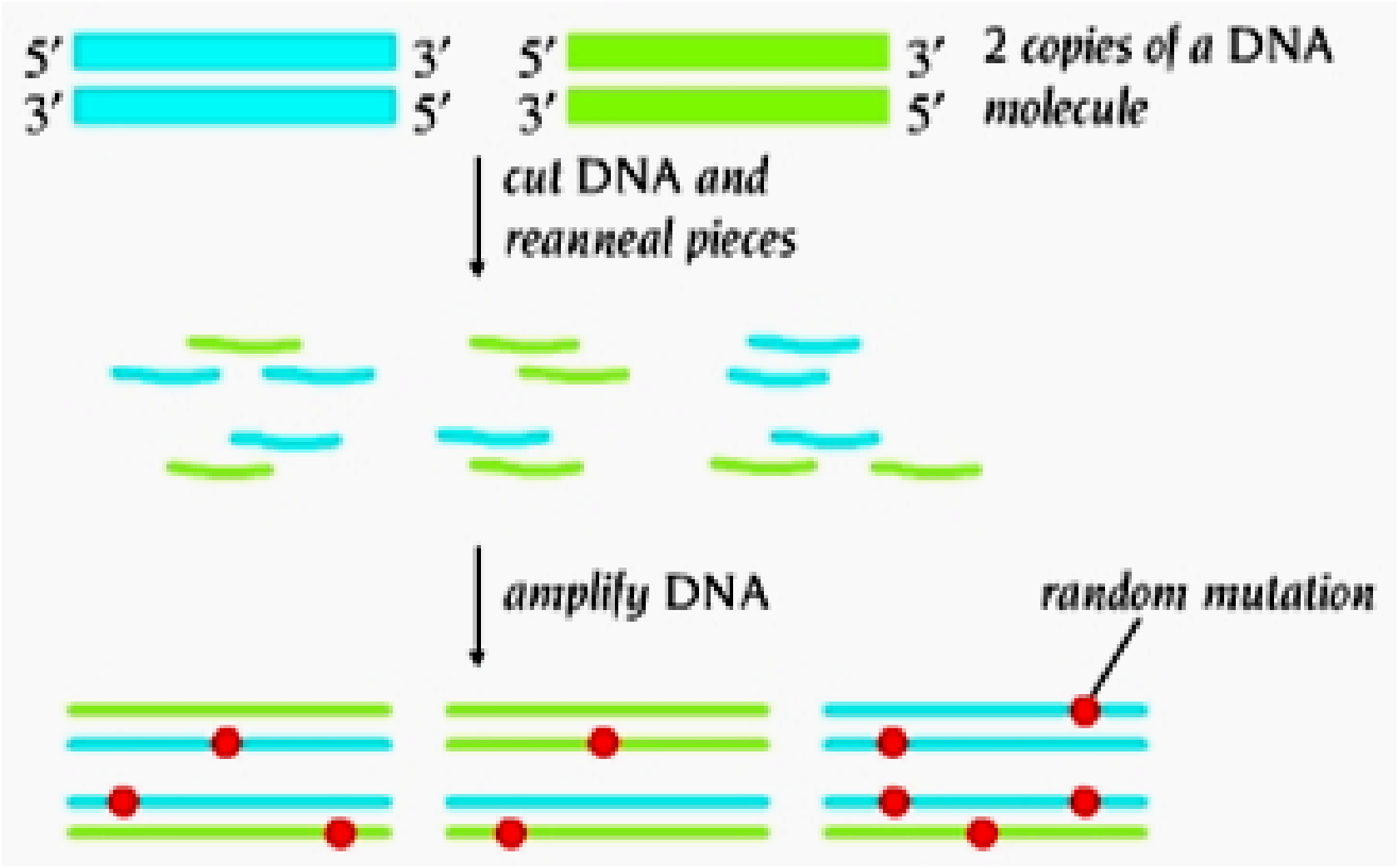

Wild-type
Random mutagenesis and screening
Sequential random mutagenesis
Sexual recombination
DNA shuffling
Random mutagenesis
Pairwise recombination
Poolwise recombination
-Useful mutations accumulate sequentially
-Many useful mutations are thrown away
-Deleterious mutations linger
-Useful mutations are combined
-Deleterious mutations are lost
-Progeny can have no more than two parents
-Many useful mutations are combined in a single cross
-Deleterious mutations are lost
-Progeny can have many parents

DNA Shuffling

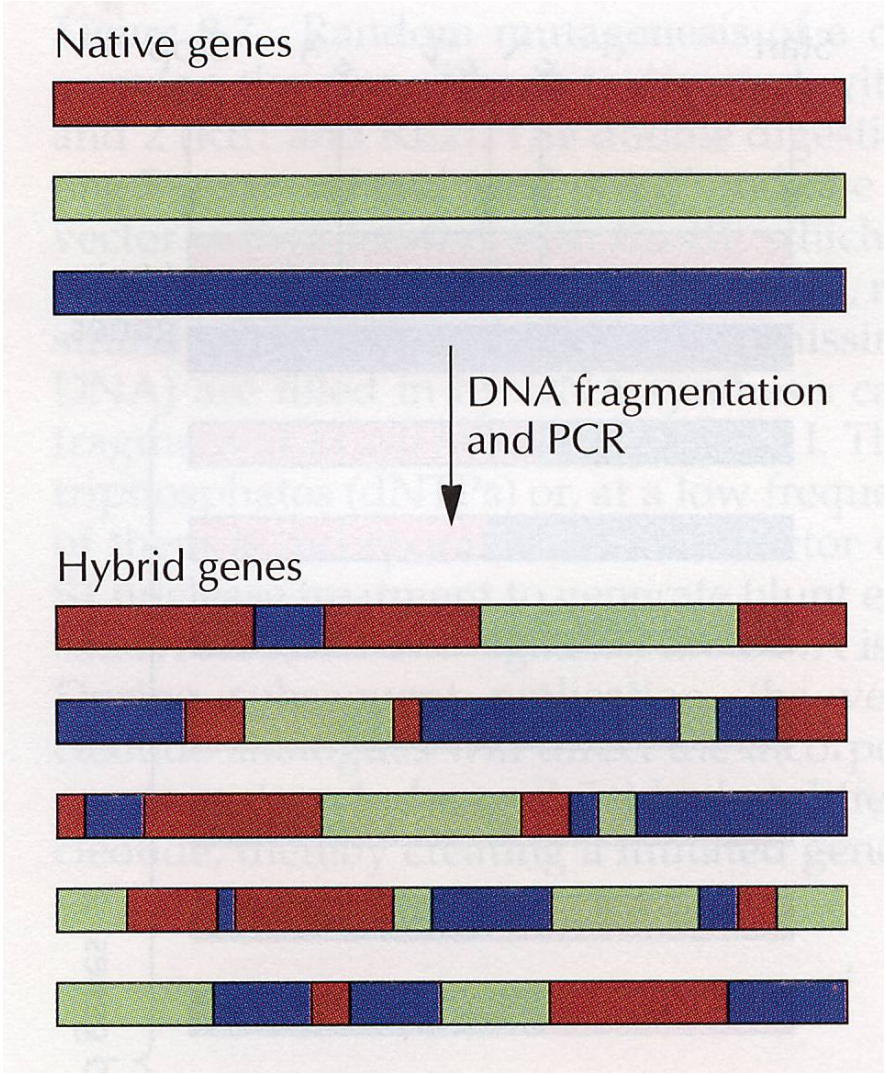


- 3.7 crossovers per 2.1 kb gene (1%) with a low mutagenesis rate (0,01%)
- successfully to recombine parents with only 63% DNA sequence identity

Gene shuffling: “sexual PCR”



Family Shuffling

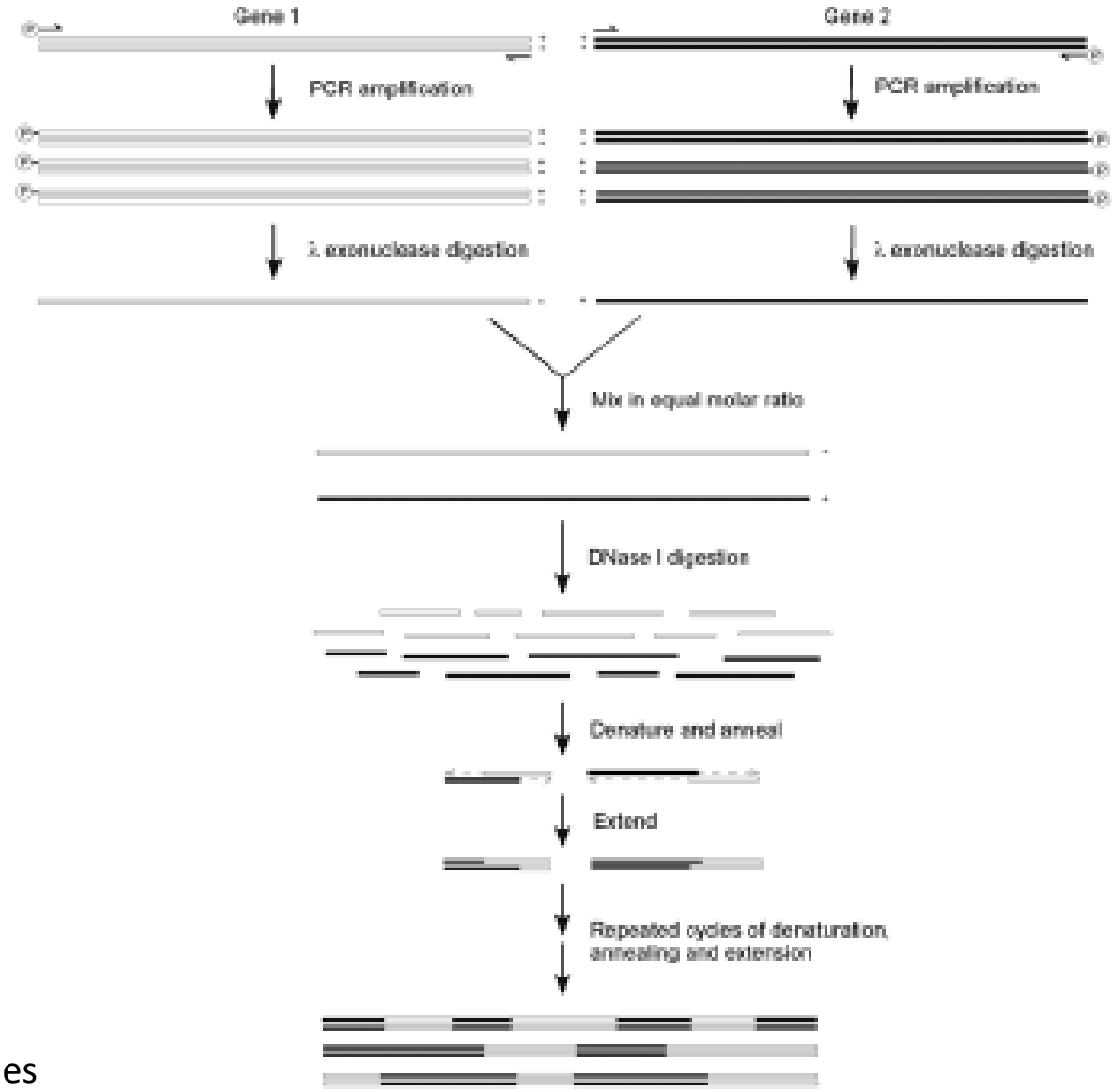


Genes coming from the same gene family -> highly homologous

-> Family shuffling

Family Shuffling with Single-Stranded DNA

generated by cloning into phagemids
or by phosporilathed DNA digestion

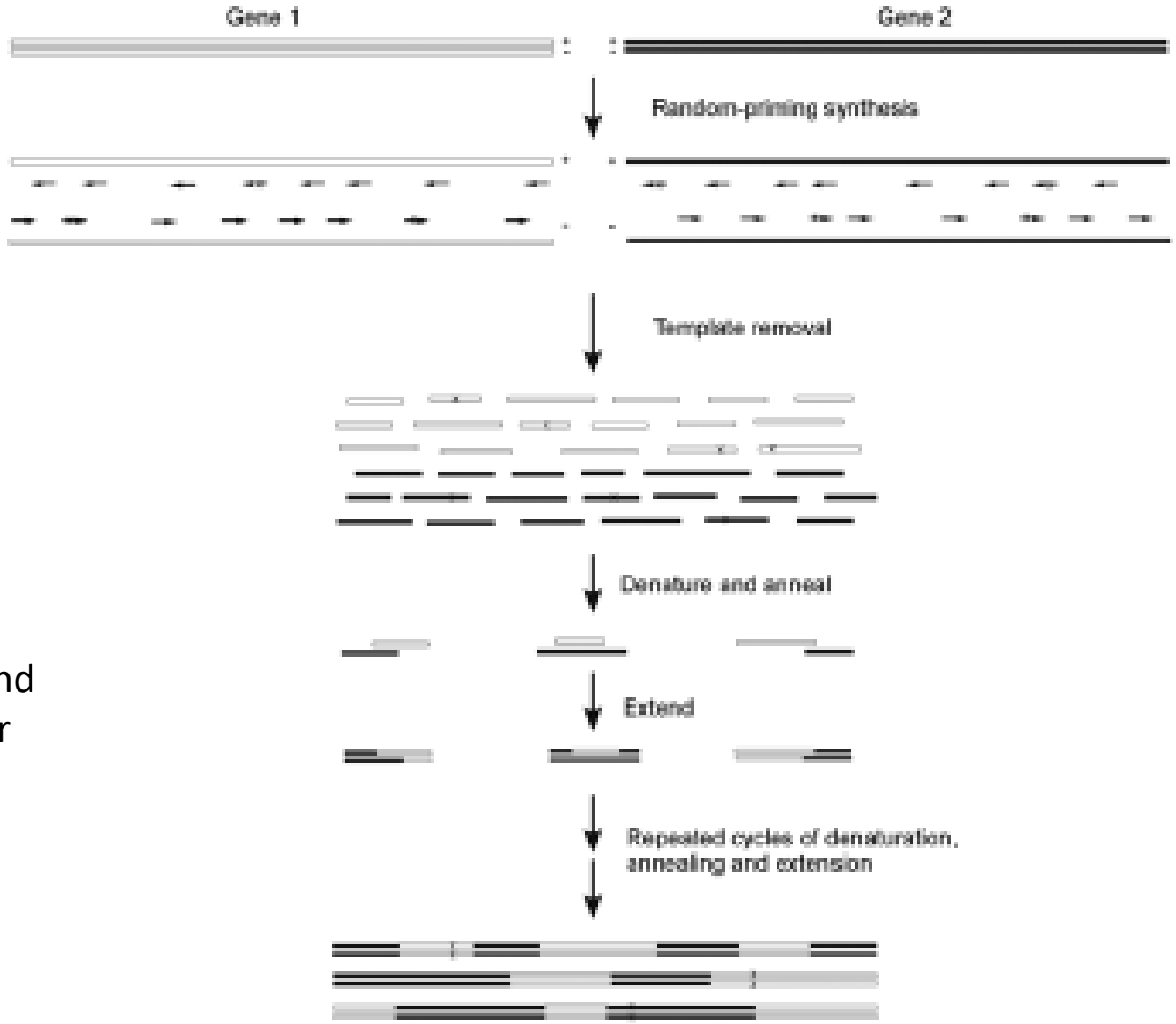


14% rate of chimeric genes

In Vitro DNA Recombination by Random Priming (RPR)

RPR has several potential advantages over DNA shuffling:

- random-priming DNA synthesis is independent of the length of the DNA template
- It can be used single-stranded DNA or RNA templates
- mutations introduced by misincorporation and mispriming can further increase the sequence diversity



Directed evolution of proteins by exon shuffling

Joost A. Kolkman and Willem P.C. Stemmer*

Figure 1. Intron and exon classes. Introns can be classified into phase 0, phase 1, and phase 2 introns depending on their position relative to the reading frame of the gene. Exons can be divided into nine different classes depending on the phases of their flanking introns. Codons are indicated in color.

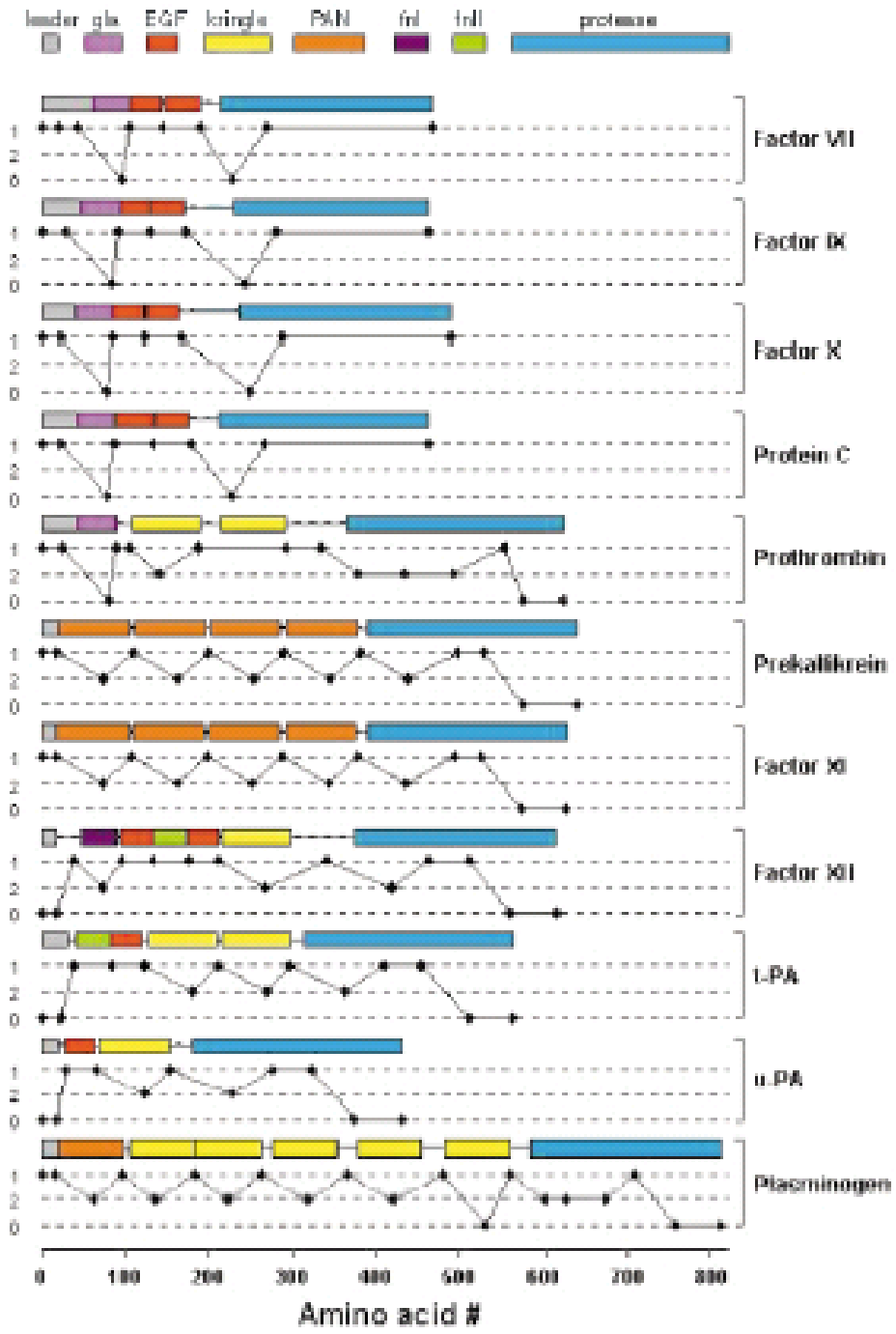
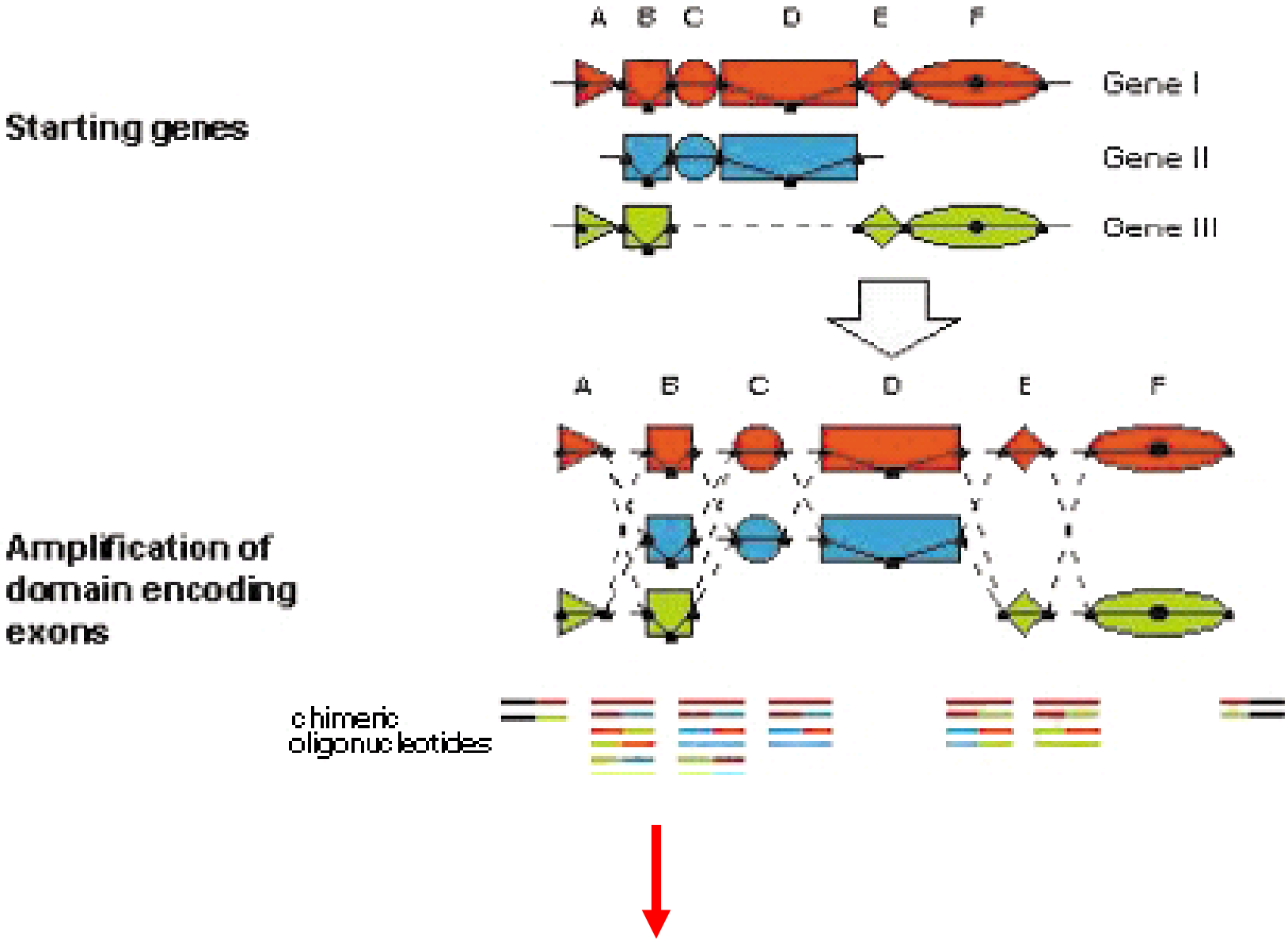


Figure 2. Domain structures and splice frame diagrams of the regulatory proteases of blood coagulation and fibrinolysis. Colors indicate different domains: *blue*, serine protease domain; *red*, EGF-like domain; *pink*, Gla domain; *gray*, pre-pro leader sequence (leader); *yellow*, kringle domain; *orange*, PAN domain; *purple*, fibronectin type II domain (fn2); *green*, fibronectin type I domain (fn1). The positions of intron-exon junctions are denoted by solid diamonds. The phase class of

Table 1. Correlation between mobile domains and the exon structure of their encoding genes

Domain	Exon class	Protein	Domain	Exon class	Protein
EGF-like domain	1-1	Epidermal growth factor precursor[21]	Thyroglobulin type I repeat	1-1	Nidogen[24]
	1-1	Tissue-type plasminogen activator[22]		1-1; 1-2-1	Thyroglobulin[25]
	1-1	Factor X[23]		1-2	Major gastrointestinal tumor-associated protein GA733-2[26]
Kringle domain	1-2-1	Hepatocyte growth factor[24]	LDL-receptor domain class A	1-1	Enteropeptidase (A. Holzinger, unpublished data)
	1-2-1	Tissue-type plasminogen activator[22]		1-1	Low density lipoprotein receptor[27]
	1-2-1; 1-1	Prothrombin[25]		No correlation	Complement component C9[28]
Fibronectin type I domain (finger domain)	1-1	Fibronectin[26]	Link domain	1-1	Proteoglycan link protein[62]
	1-1	Factor XII[27]		1-2-1	CD 44 antigen[63]
	1-1	Tissue-type plasminogen activator[22]		1-2-1; 1-1	Versican core protein[64]
Fibronectin type II domain	1-1	Fibronectin[26]	Thrombospondin type 1 domain (TSP repeat)	1-1	Thrombospondin 1[65]
	1-1	Macrophage mannose receptor[28]		1-1; 2-1	Properdin[61]
	1-2-1	Factor XII[27]		No correlation	Complement component C9[28]
Fibronectin type III domain	1-1-1; 1-1; 1-0-1	Fibronectin[26]	Immunoglobulin-like domain	1-1	PECAM-1[59]
	1-1-1; 1-0-1	Axonin-1[29]		1-1-1; 1-2-1	Neural cell adhesion molecule L1[60]
	1-0-1; 1-1-0-1	Kallmann syndrome protein[31]		1-2-1; 1-1-1;1-0-1	Axonin-1[29]
PAN domain	1-2-1	Plasminogen[32]	C-type lectin domain	1-1	L-selectin[65], E-selectin[66], P-selectin[28]
	1-2-1	Prekallikrein[33]		1-1-0-1	Versican core protein[64]
	1-2-1	Factor XI[34]		1-0-2-x	Kupffer cell receptor[67]
Gla domain	1-0-1	Factor X[23]	MAM domain	1-1-0-1	Neuropilin-1[53]
	1-0-1	Prothrombin[25]		1-1-1-1	Meprin A α- and β-subunit[68]
	1-0-1	Protein S[35,36,37]		1-1-2-0-1	Enteropeptidase (A. Holzinger, unpublished data)
Sushi domain (SCR repeat, CCP module)	1-1	Factor XIIIb[38]	von Willebrand factor type A domain	1-0-1; 1-1-1	Cartilage matrix protein[69]
	1-1	P-selectin[39]		1-0; 1-1-1	Collagen α2(VI) chain[70]
	1-1-1	Haptoglobin[40]		1-0-2-0-1	Integrin α-X[71]
SRCR domain	1-2-2-1	Enteropeptidase (A. Holzinger, unpublished data)	Somatomedin B domain	1-1	Vitronectin[72]
	1-1	T-cell surface glycoprotein CD5 (O. Padilla, unpublished data)		1-1	Plasma-cell membrane glycoprotein PC-1 (M. Bozzali, unpublished data)
	1-2-1	Complement factor I (unpublished data)		1-1	Placental protein 11 (K. C. Worley, unpublished data)
Kunitz/Bovine pancreatic trypsin inhibitor type	1-1	Tissue factor pathway inhibitor[40]	WAP-type four-disulfide core	1-1	Whey acidic protein[73]

In Vitro Exon Shuffling



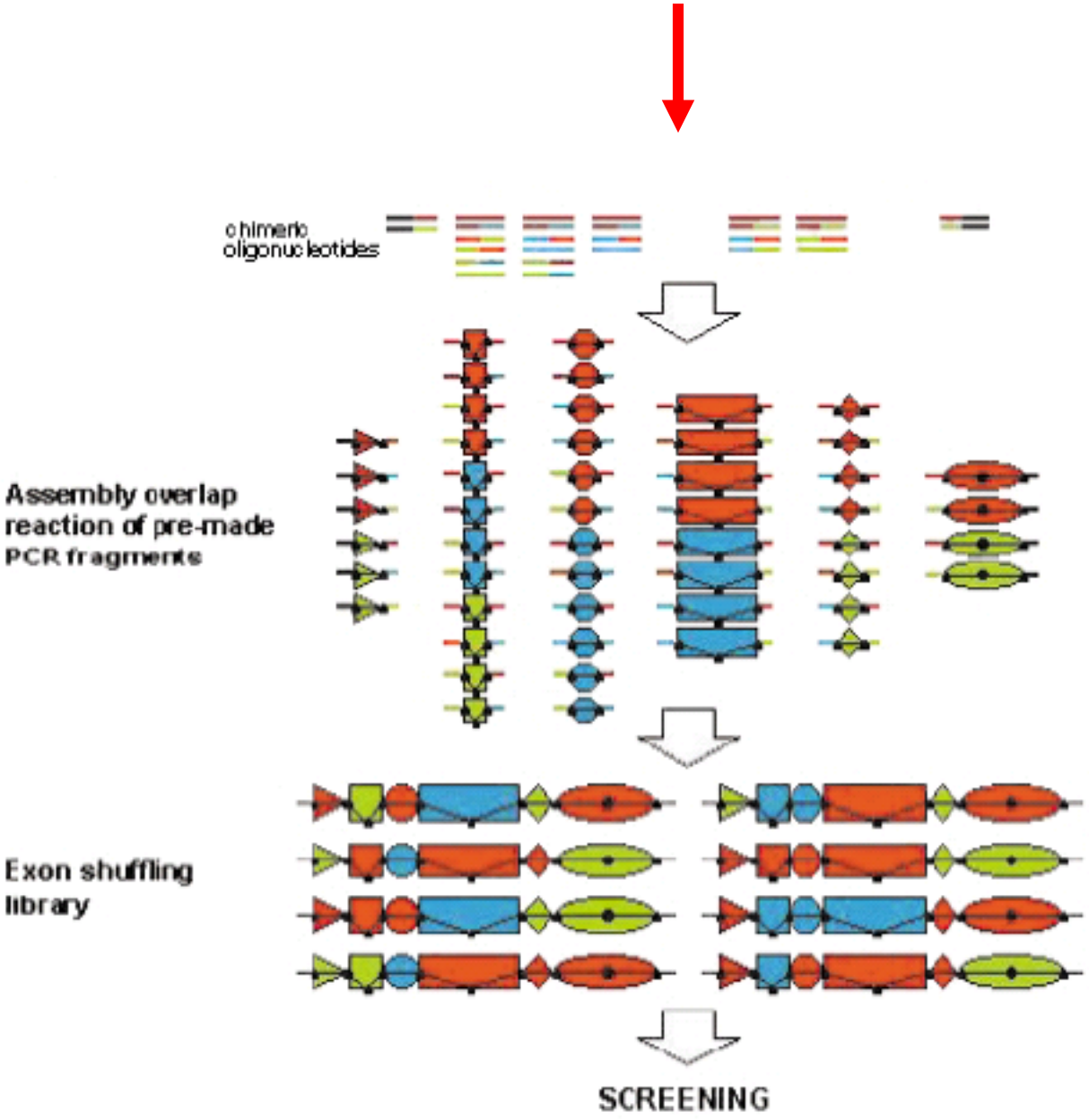

chimeric
oligonucleotides
Assembly overlap
reaction of pre-made
PCR fragments
Exon shuffling
library
SCREENING

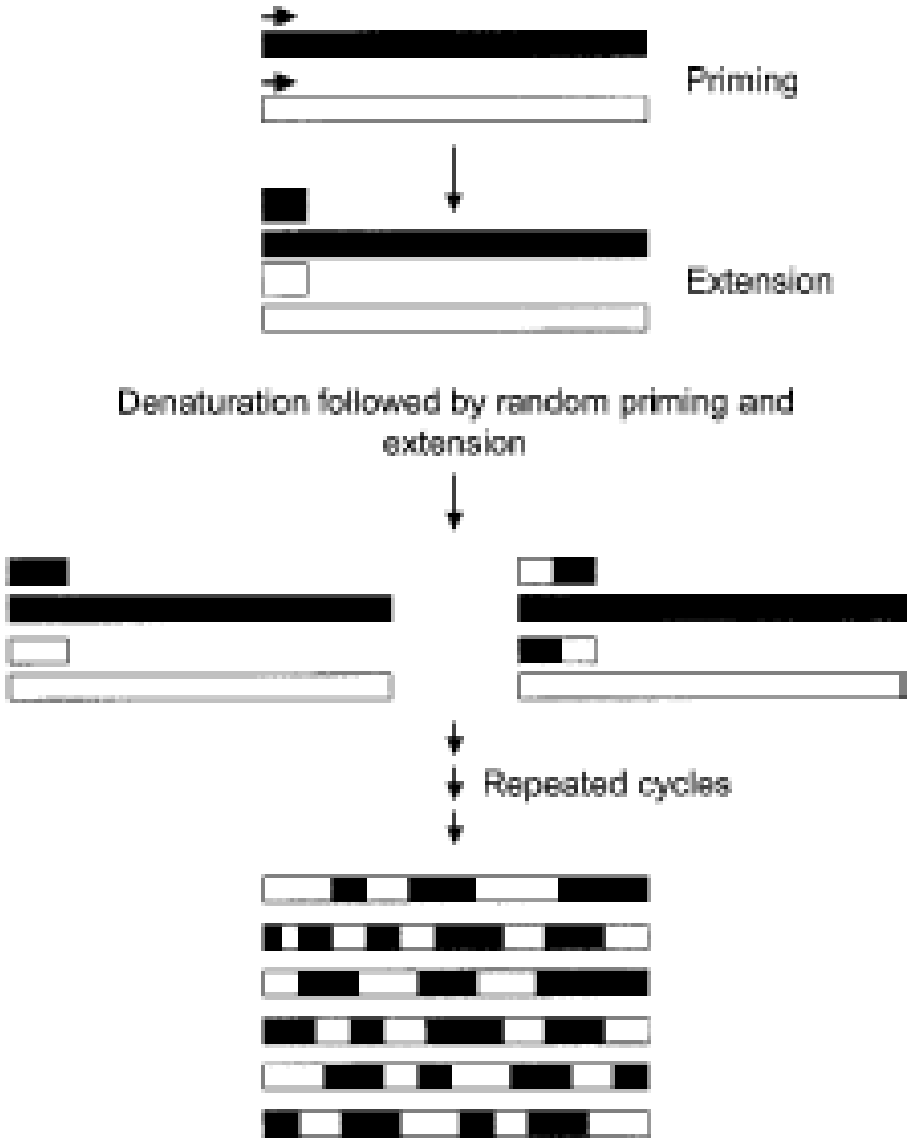
Table 2. Examples of library formats based on exon shuffling

Orthologous exon shuffling	Substitution of exons with equivalent exons from the same gene of different species
Paralogous exon shuffling	Substitution of exons with homologous exons from different genes of the same species
Orthologous domain shuffling	Substitution of domains with equivalent domains encoded by the same gene of different species
Paralogous domain shuffling	Substitution of domains with homologs encoded by different genes of the same species
Domain number variation	Deletion, duplication or insertion of protein domains
Functional homolog shuffling	Substitution of domains with non-homologous domains that are functionally related
De novo protein assembly	Assembly of multiple, independent domains into novel proteins

Staggered Extension Process (StEP) In Vitro Recombination

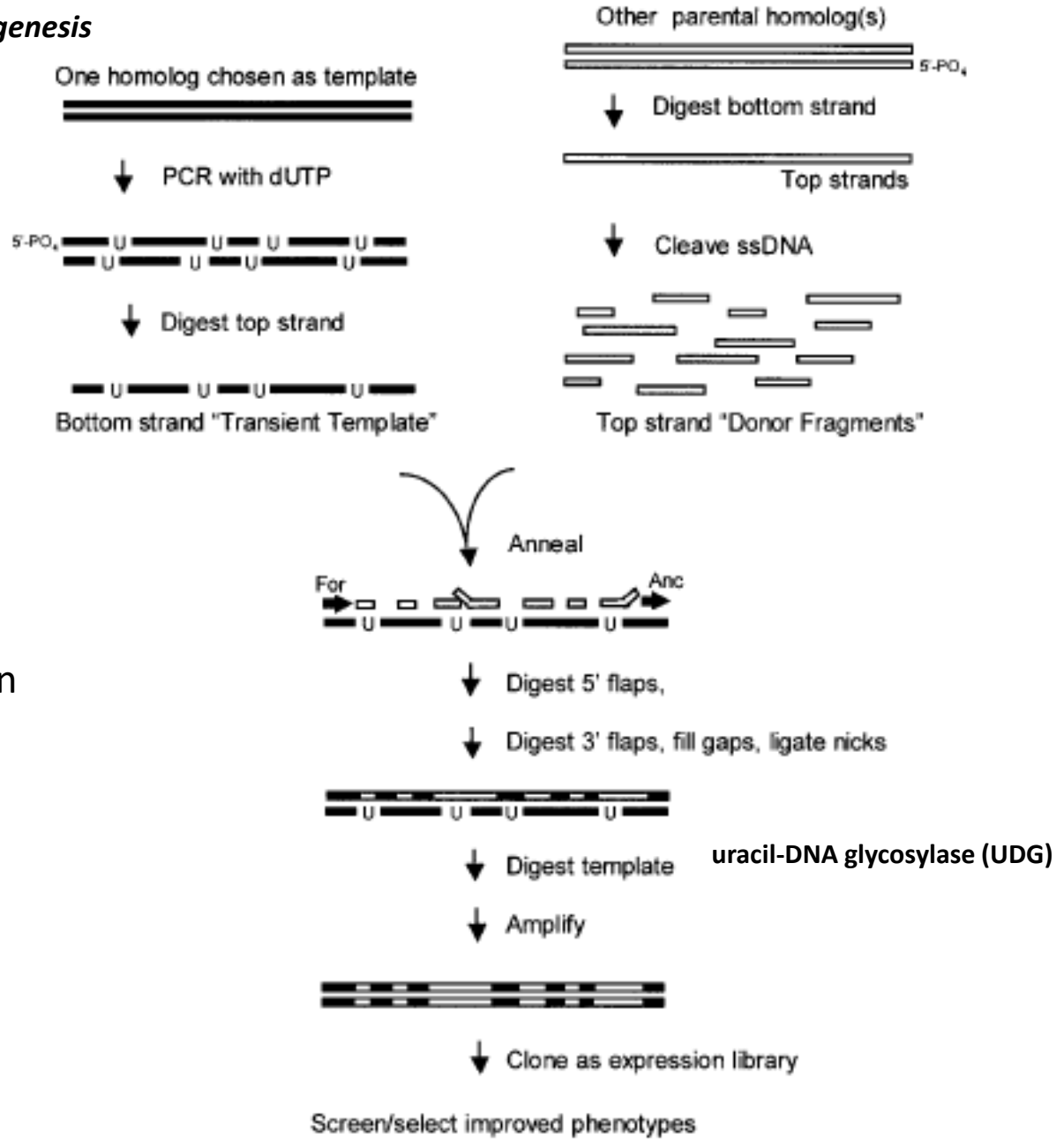
based on cross hybridization of growing gene fragments during polymerase-catalyzed primer extension

It is much easier than DNA shuffling



RACHITT

Gene Family Shuffling by Random Chimeragenesis on Transient Templates



average 12 crossovers per gene in a single round

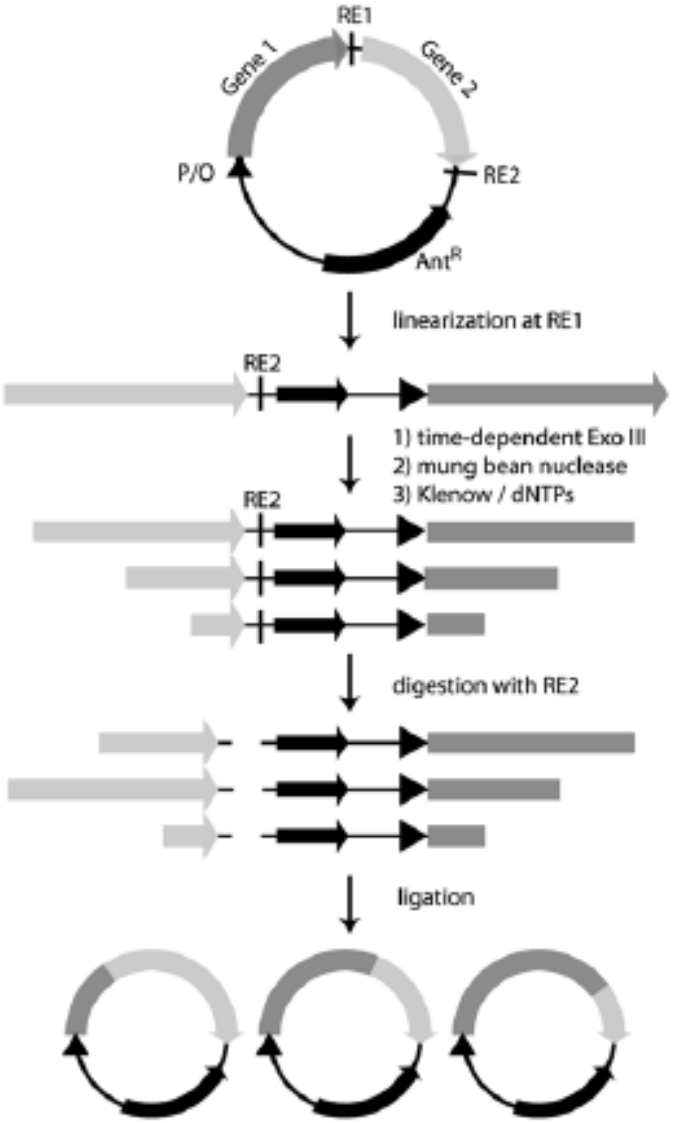
The Creation of ITCHY Hybrid Protein Libraries

An ITCHY library created from a single gene consists of genes with internal deletions and duplications.

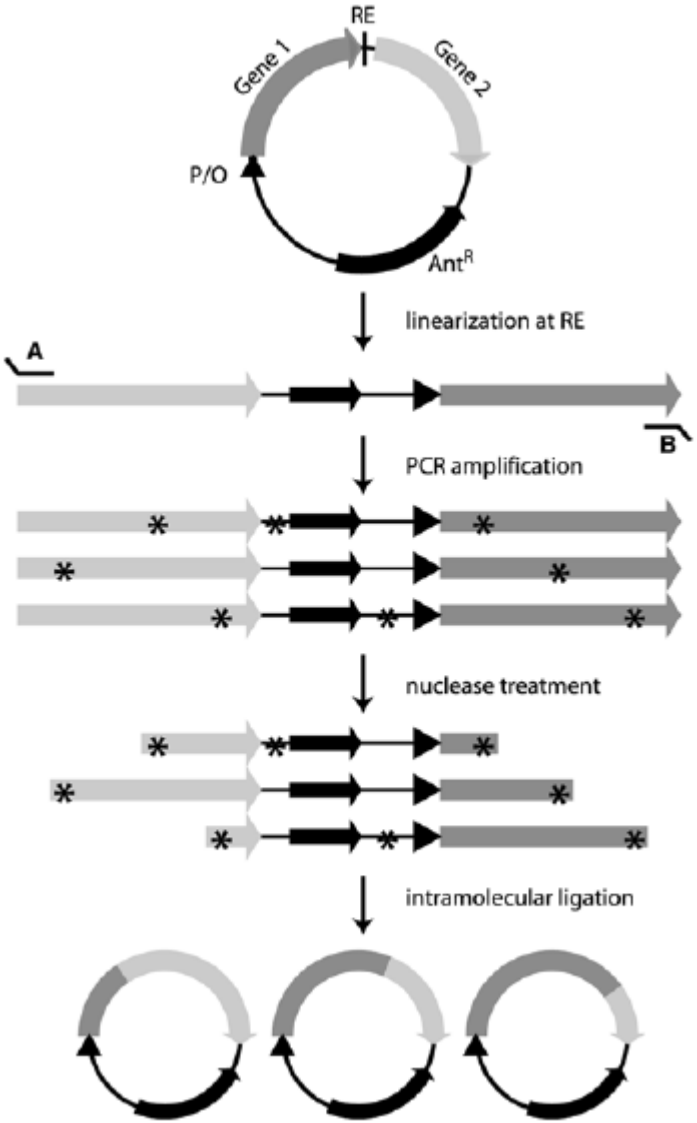
An ITCHY library created between two different genes consists of gene fusions created in a DNA-homology independent fashion.

The Creation of ITCHY Hybrid Protein Libraries

Using α-phosphorothioate dNTPs



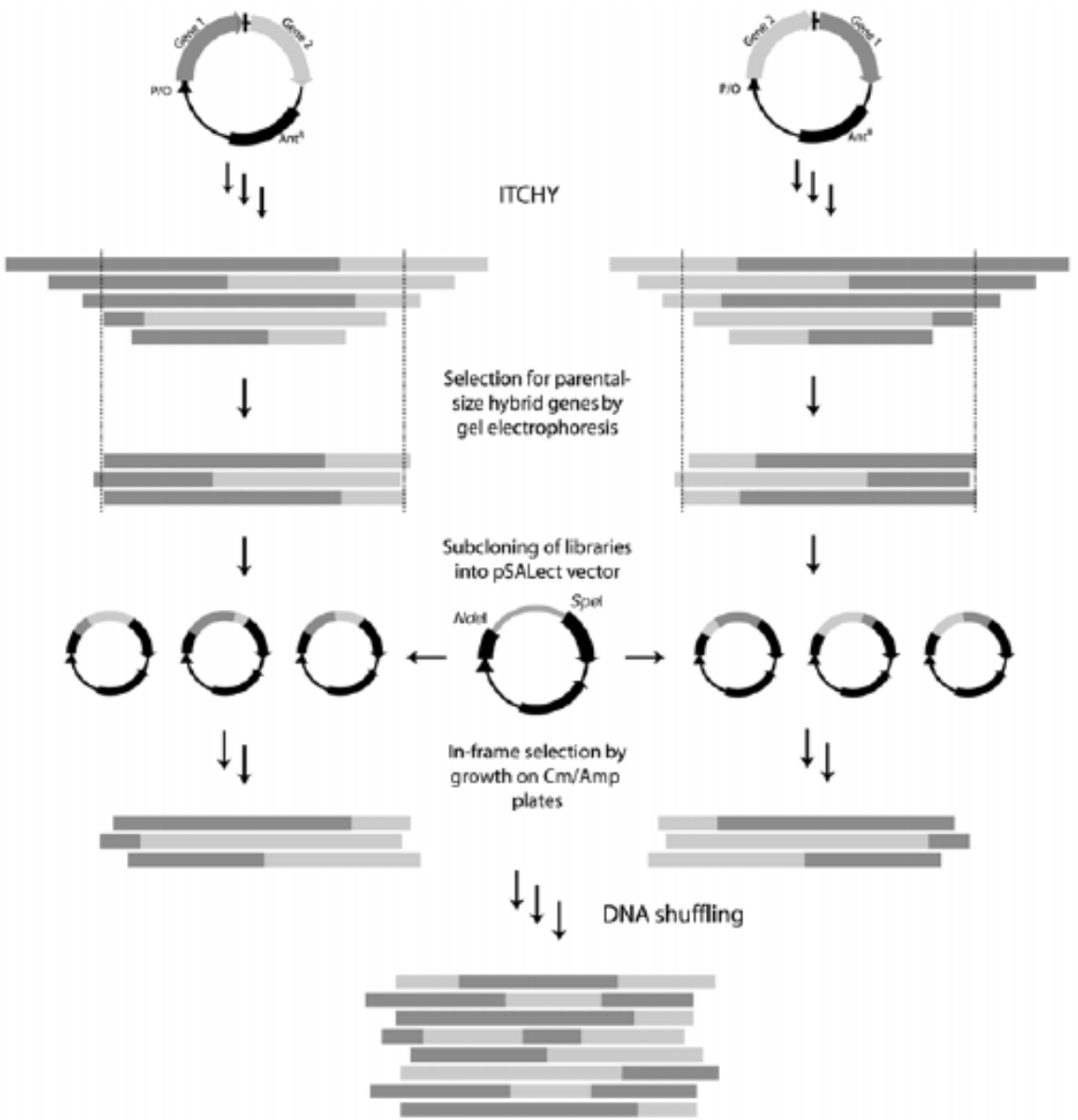
Blunt ends generation



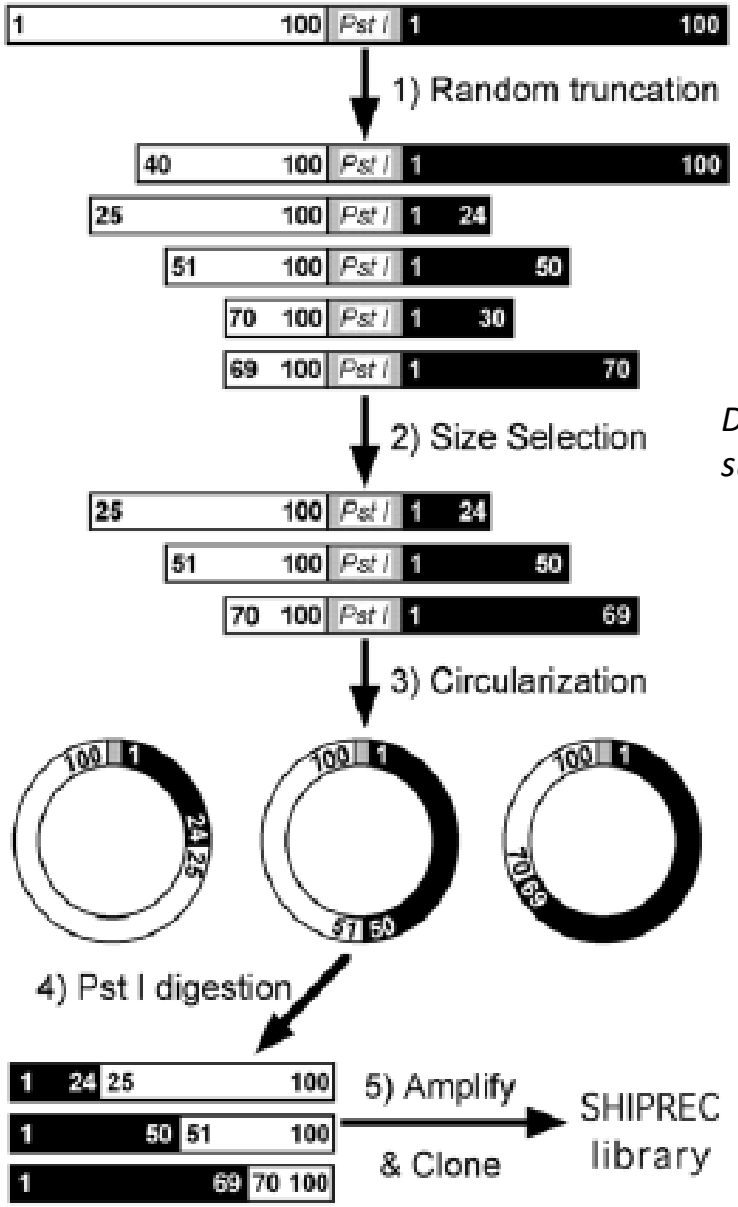
Preparation of SCRATCHY Hybrid Protein Libraries

Size- and In-Frame Selection of Nucleic Acid Sequences

A combination between ITCHY and DNA shuffling



Sequence Homology-Independent Protein Recombination (SHIPREC)



DNase I and S1 nuclease

DNA corresponding to the length of the parental genes is isolated and subsequently circularized

crossovers occur at structurally related sites

to generate all possible single-crossover chimeras, SHIPREC must be performed twice starting with both possible parental gene fusions, e.g., A–B and B–A.