



CEITEC

Central European Institute of Technology

BRNO | CZECH REPUBLIC

Protein Engineering

Pavel Mazura

CEITEC – MENDELU

Department of molecular biology and radiobiology

Scaled Biosystems team

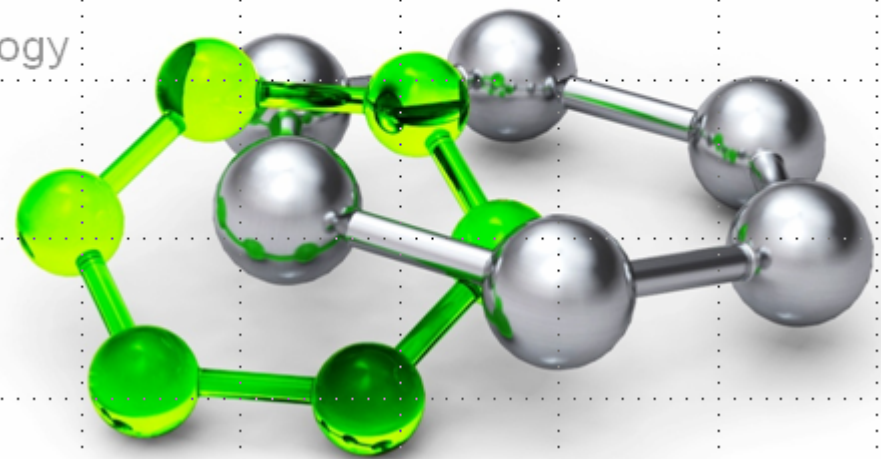
mazura@sci.muni.cz



EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE



OP Research and
Development for Innovation

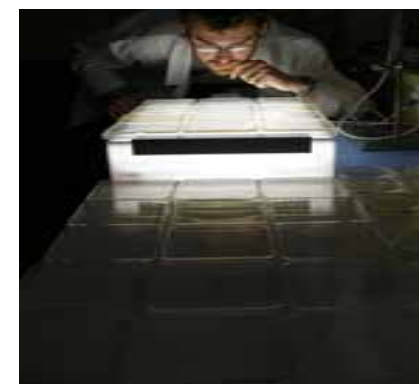


SCALED BIOSYSTEMS

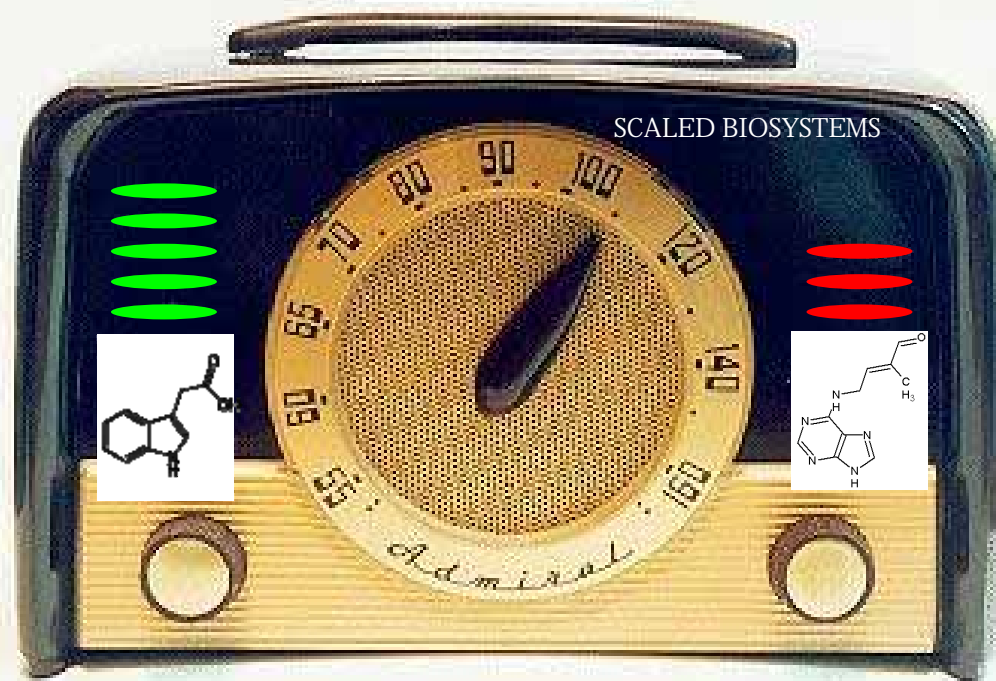
Research & Development of molecular systems for control of plant growth and morphogenesis



Dr. Tomáš Filipi



Dr. Pavel Mazura



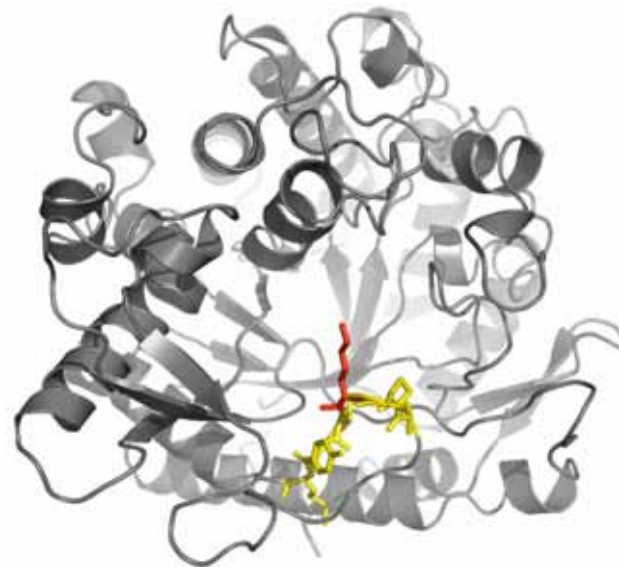
mazura@sci.muni.cz

SCALED BIOSYSTEMS

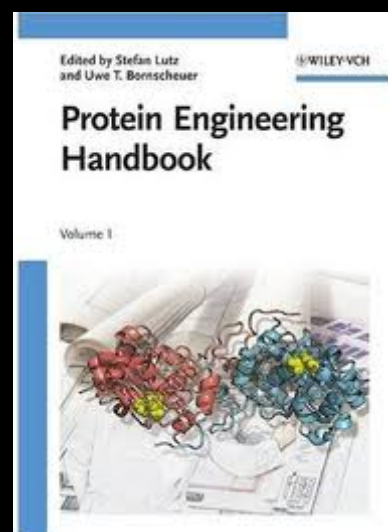
Research & Development of molecular systems for control of plant growth and morphogenesis

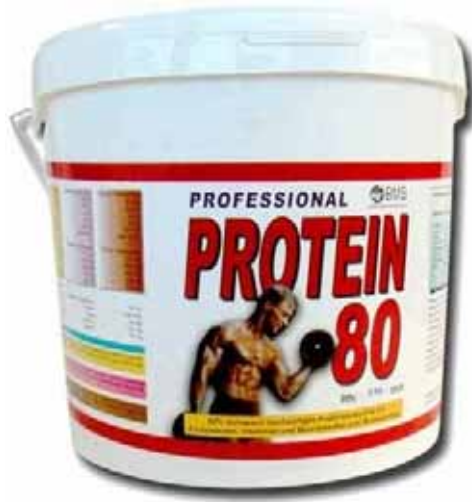
Screening and characterization of novel β -glucosidases
for target probing of plant hormone system

Bioinformatics, *E.coli* transformation and screening,
DNA isolation + (basic molecular biology techniques)

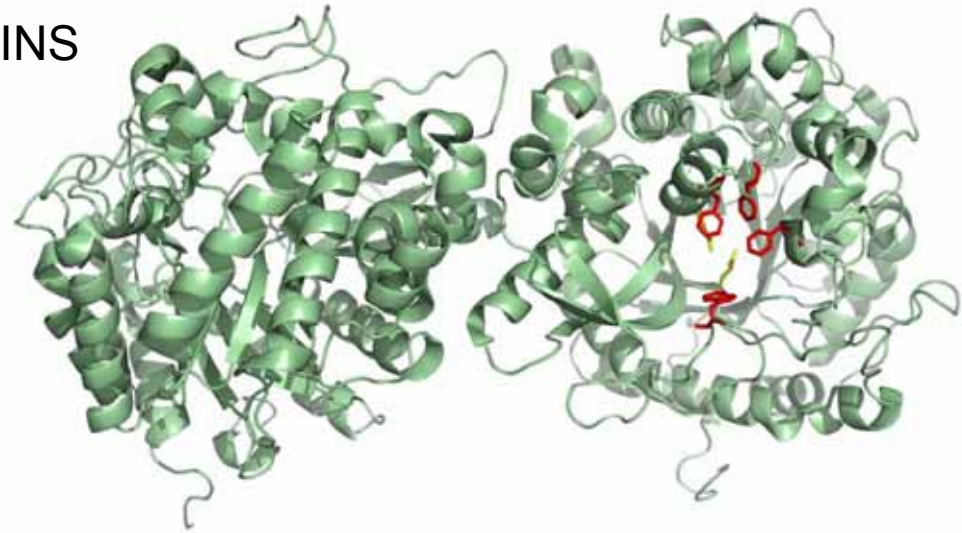


mazura@sci.muni.cz





PROTEINS



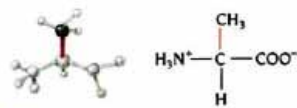
Highly complex substance that is present in all living organisms.

Proteins are of great nutritional value and are directly involved in the chemical processes essential for life.

Proteins are species-specific; They are also organ-specific; for instance, within a single organism, muscle proteins differ from those of the brain and liver.
(Encyklopedia Britannica)

Hydrophobic amino acids

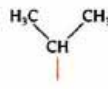
(a) Hydrophobic amino acids



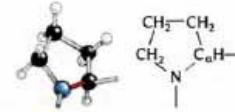
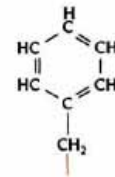
A Ala, Alanine



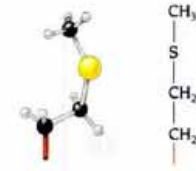
V Val, Valine



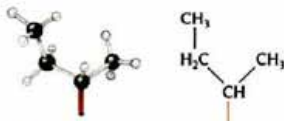
F Phe, Phenylalanine



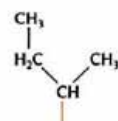
P Pro, Proline



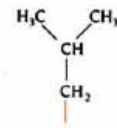
M Met, Methionine



I Ile, Isoleucine

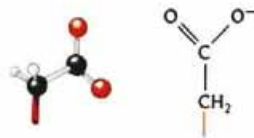


L Leu, Leucine

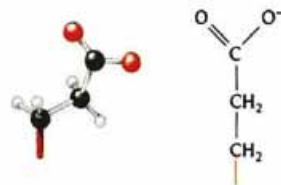


Charged amino acids

(b) Charged amino acids



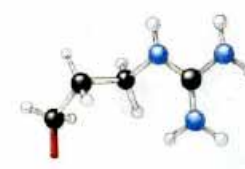
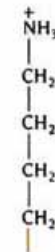
D Asp, Aspartic acid



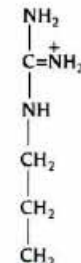
E Glu, Glutamic acid



K Lys, Lysine

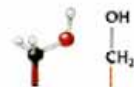


R Arg, Arginine



Polar amino acids

(a) Polar amino acids



Ser, Serine



Thr, Threonine



Tyr, Tyrosine



His, Histidine



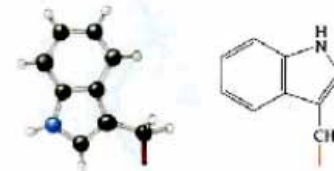
Cys, Cysteine



Asn, Asparagine

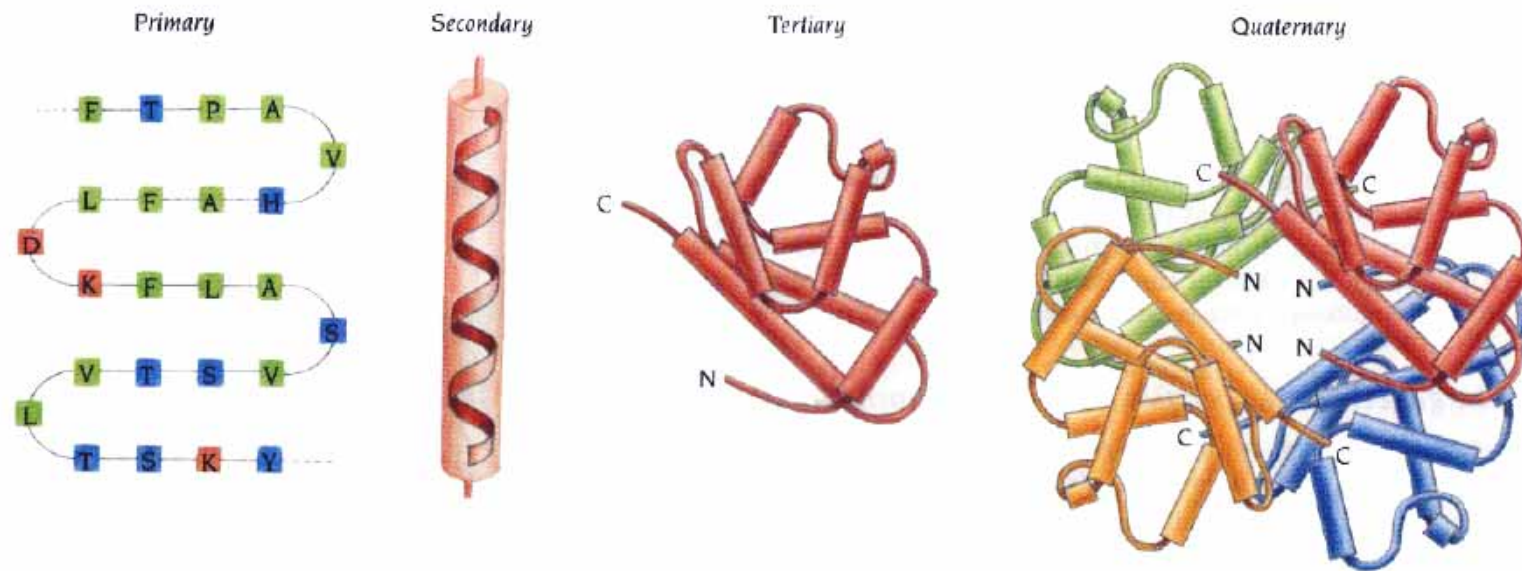


Gln, Glutamine

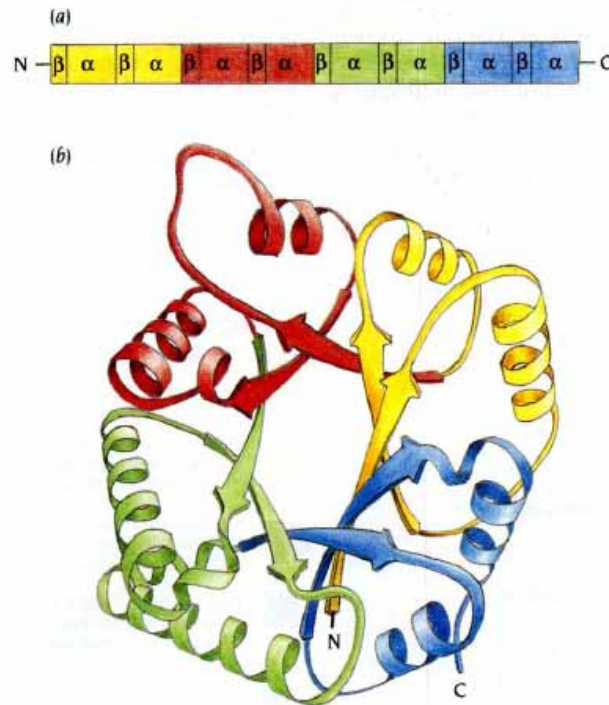


Trp, Tryptophan

Protein structure



Domains are build from structural motifs



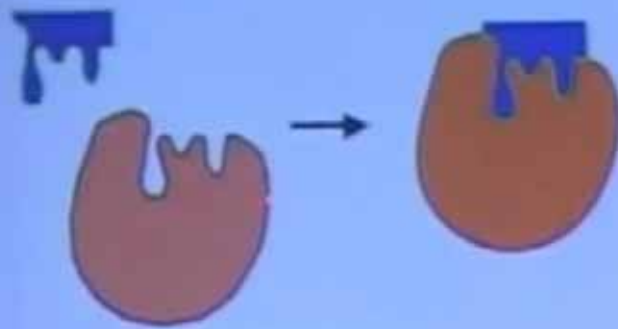
To a first approximation a polypeptide chain can be considered as a sequential arrangement of simple motifs that are formed from consecutive regions of the primary structure. For example, triosephosphate isomerase is built up from four β - α - β - α motifs that are consecutive both in the amino acid sequence (a) and in the three-dimensional structure (b).

The number of such combinations found in proteins is limited, and some combinations seem to be structurally favored.

Thus similar domain structures frequently occur in different proteins with different functions and with completely different amino acid sequences.

Proteins - the traditional view

The Lock & Key Model



one sequence =
one structure =
one function



Emil Fischer
1852-1919

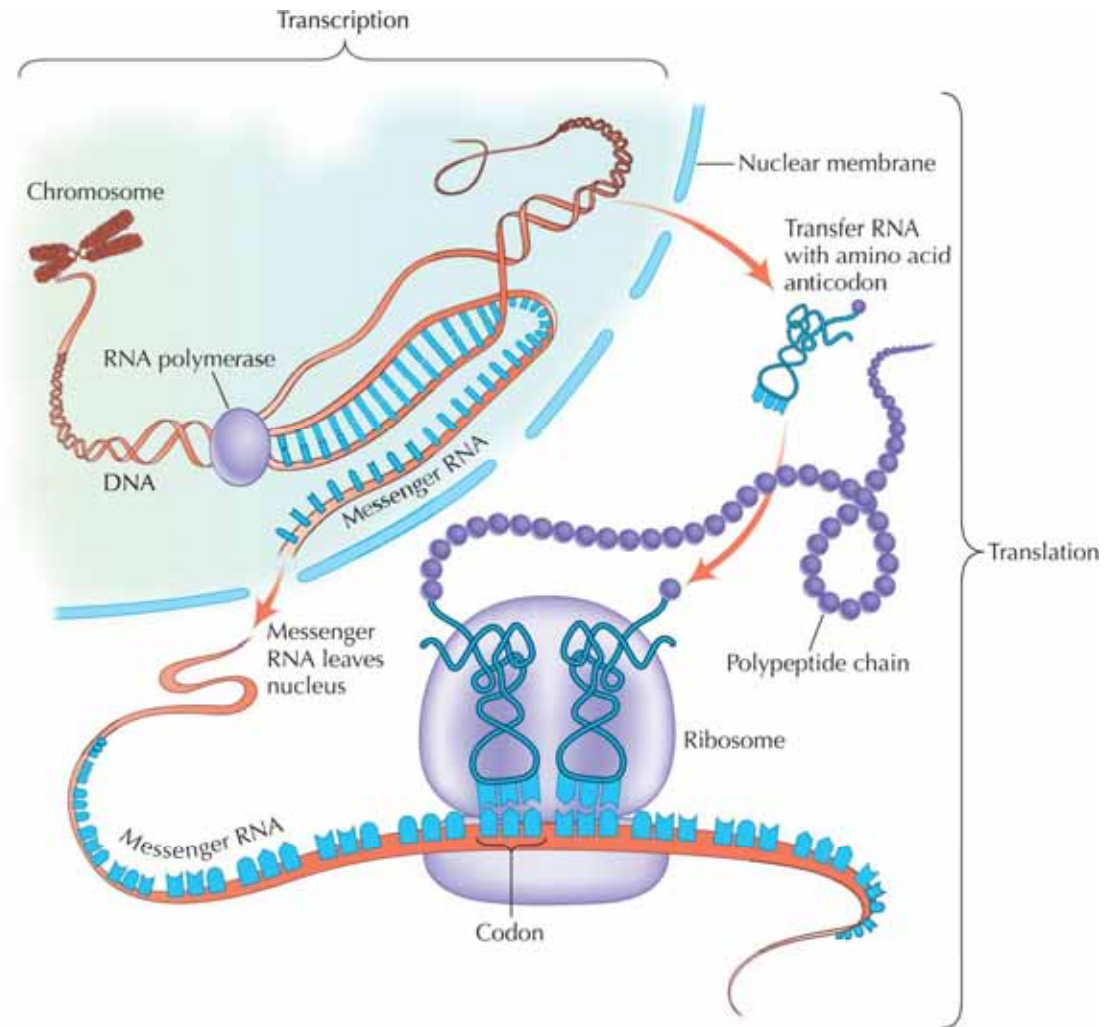
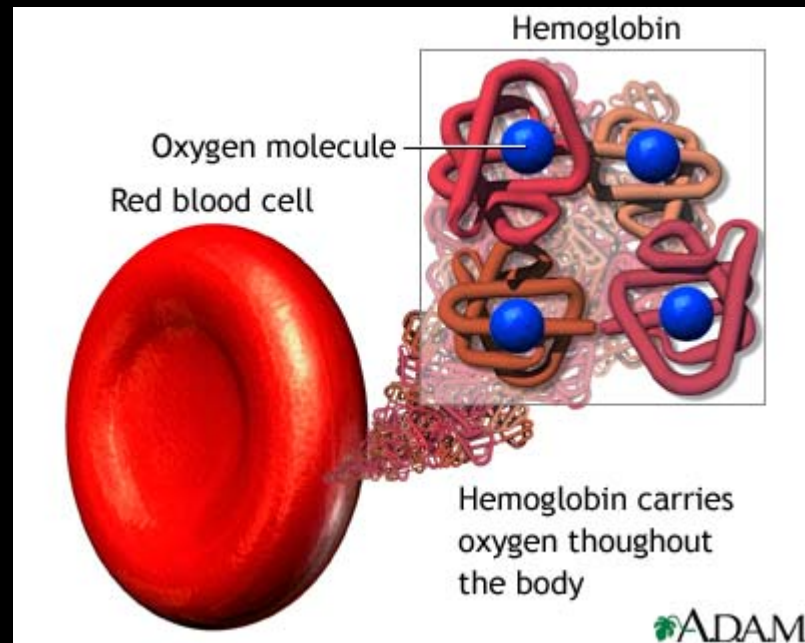
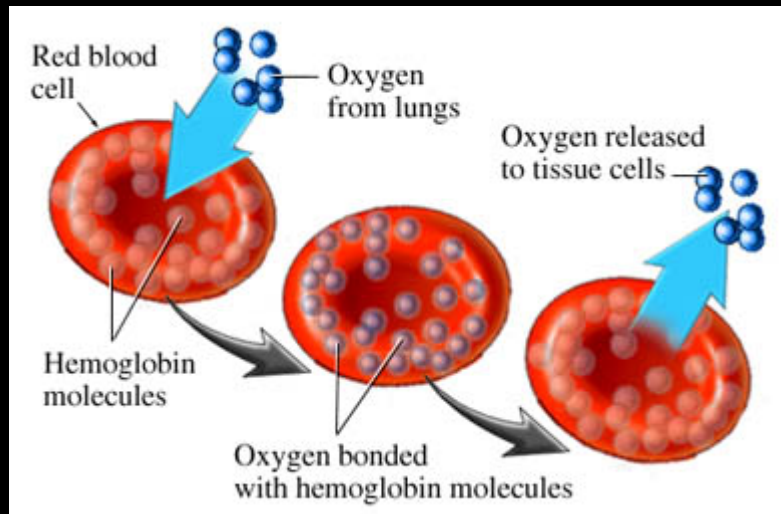
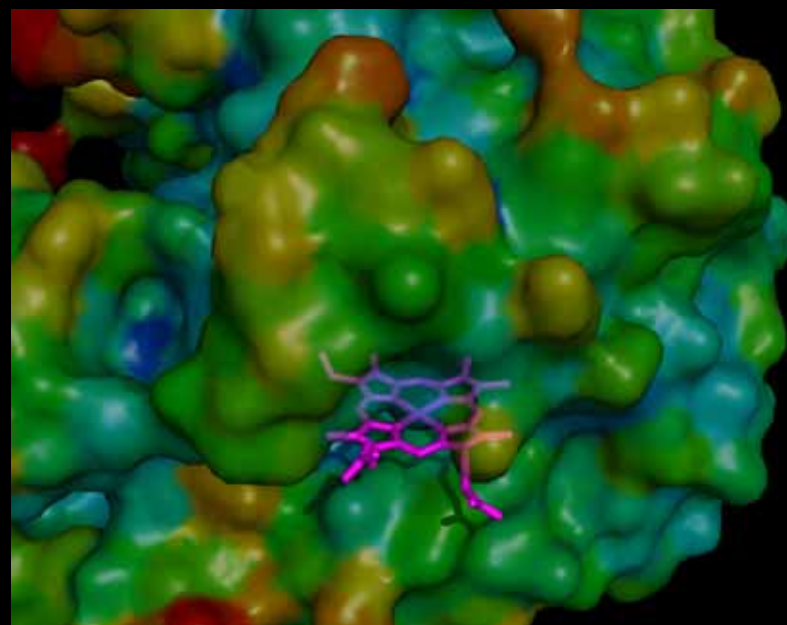
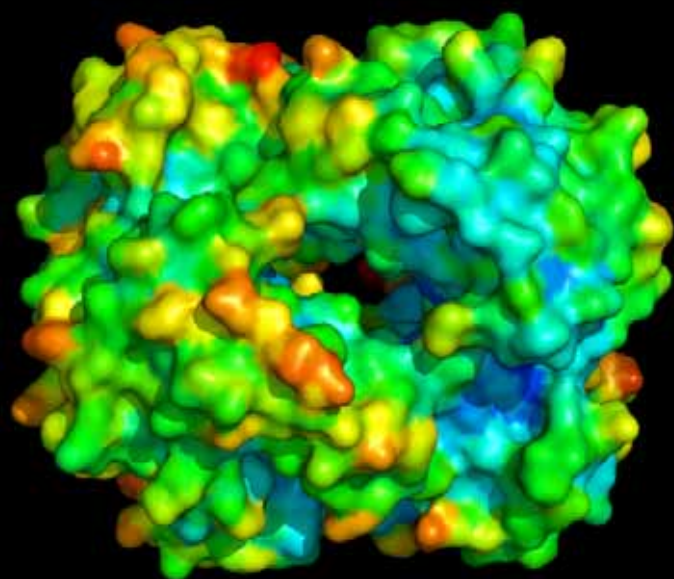


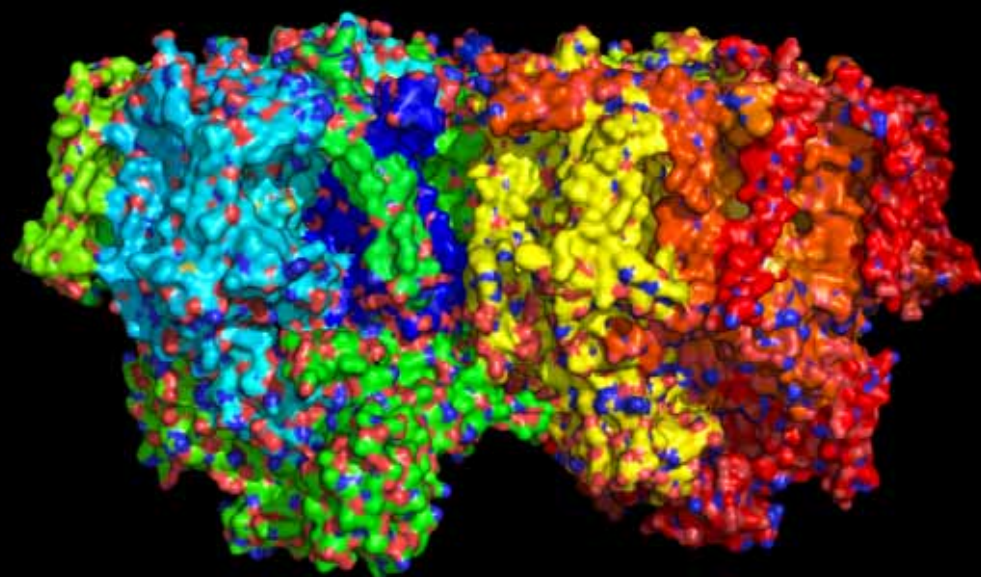
FIGURE 2.27. General mechanism of eukaryotic protein synthesis. The major steps include transcription of the DNA gene sequence into the messenger RNA template in the nucleus of the cell, translation of the DNA codons of that gene into amino acids, and their assembly into polypeptides in the cytoplasm. Important mediators of this process include transfer RNAs, splicing elements, and ribosomes.

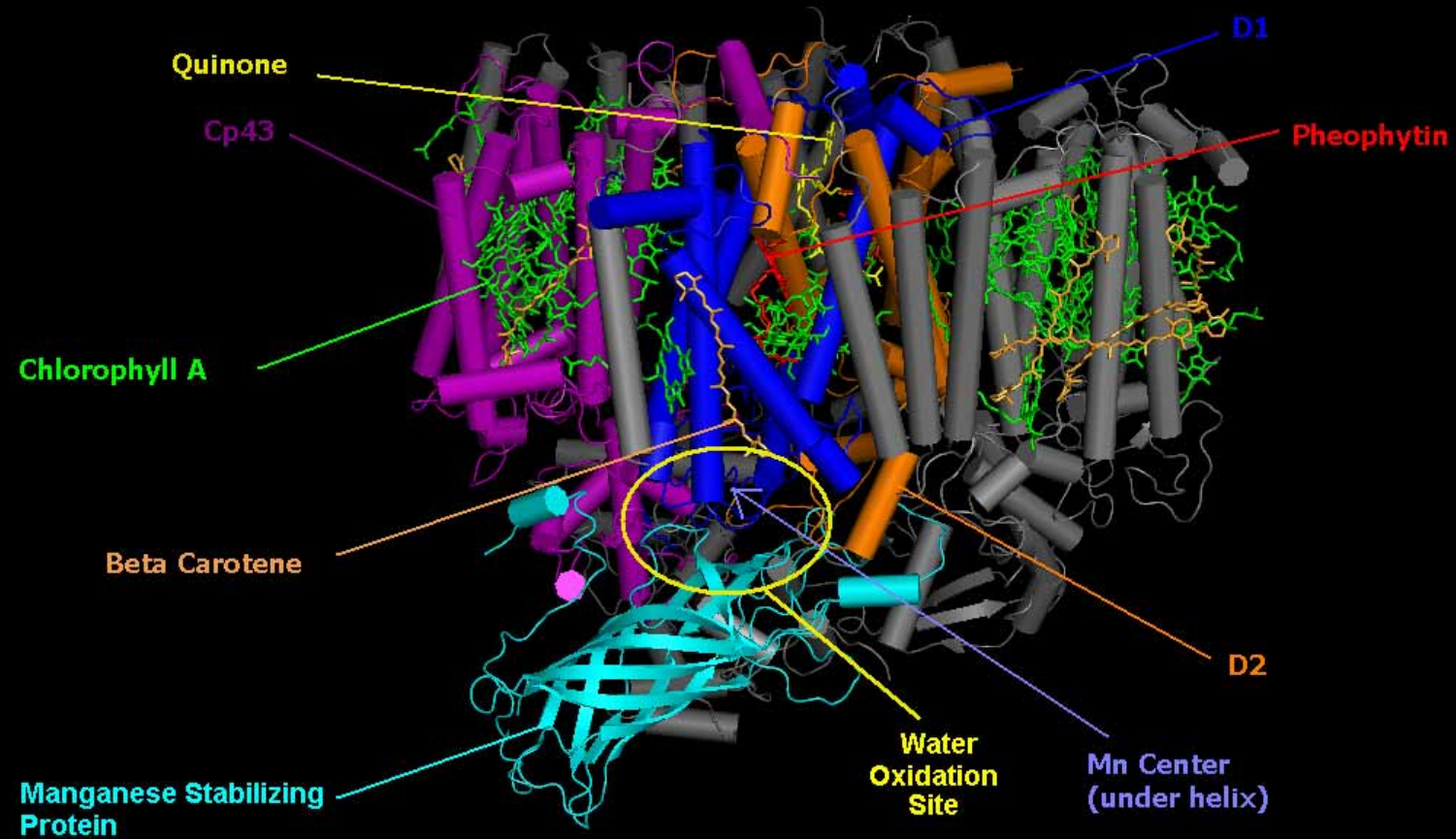
Protein Engineering

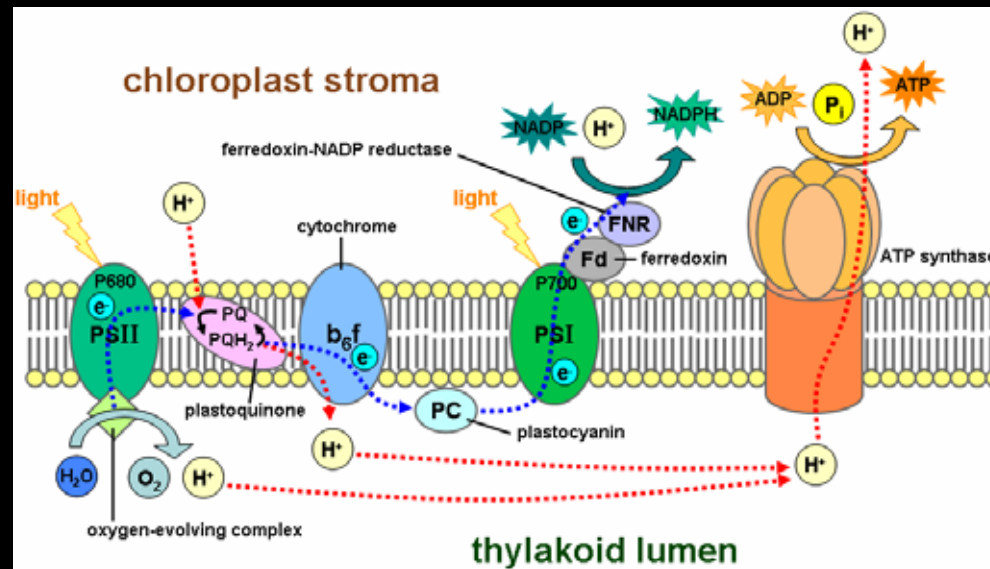
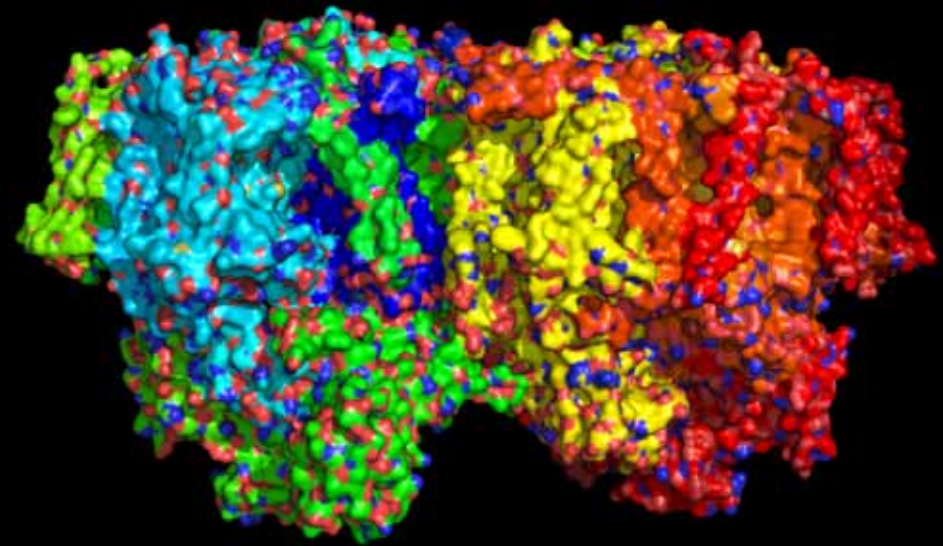
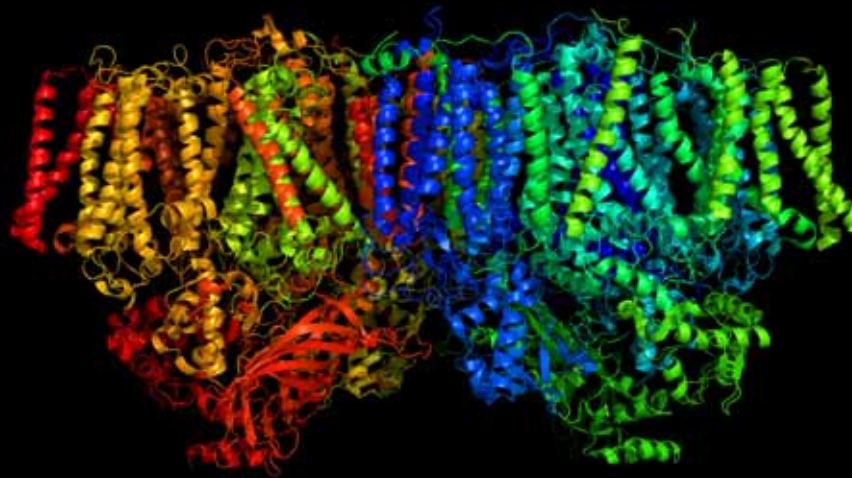
the **Big** picture

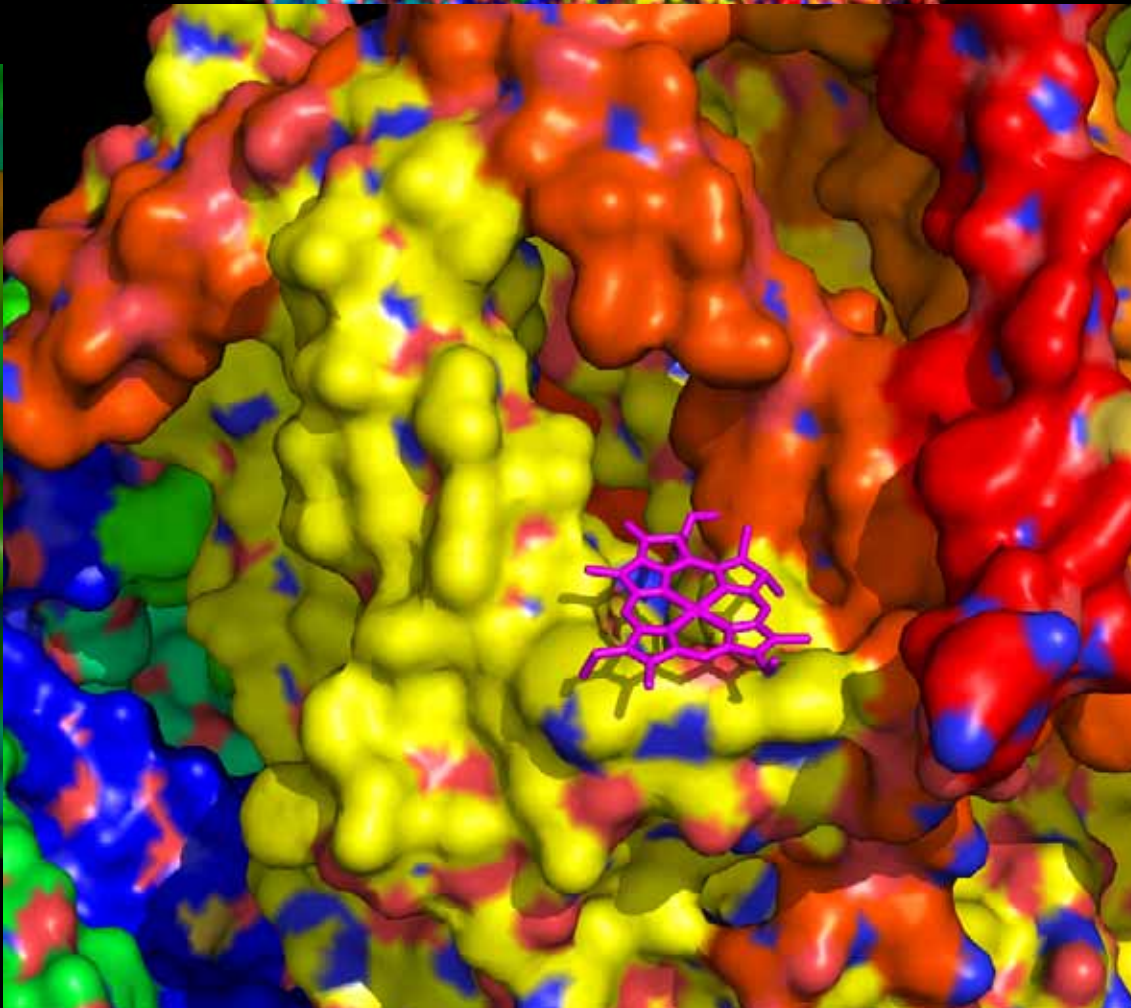
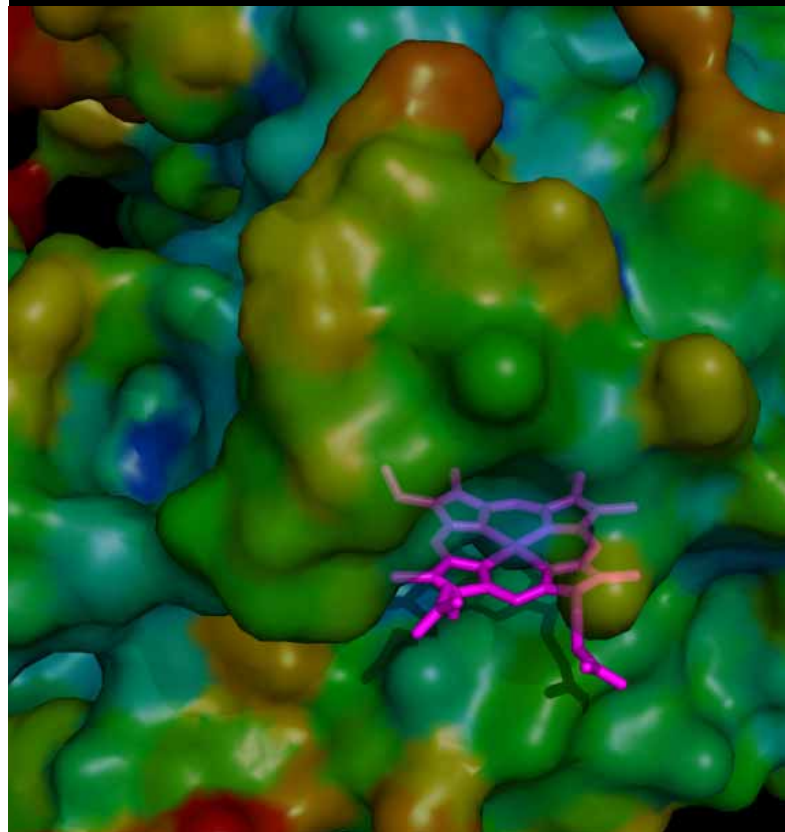
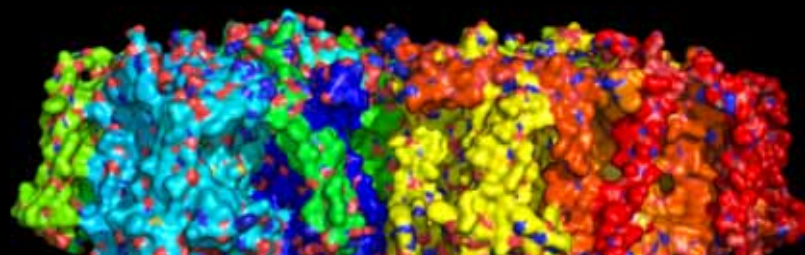
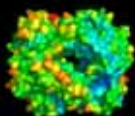




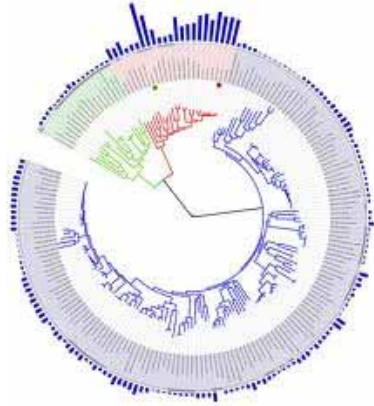




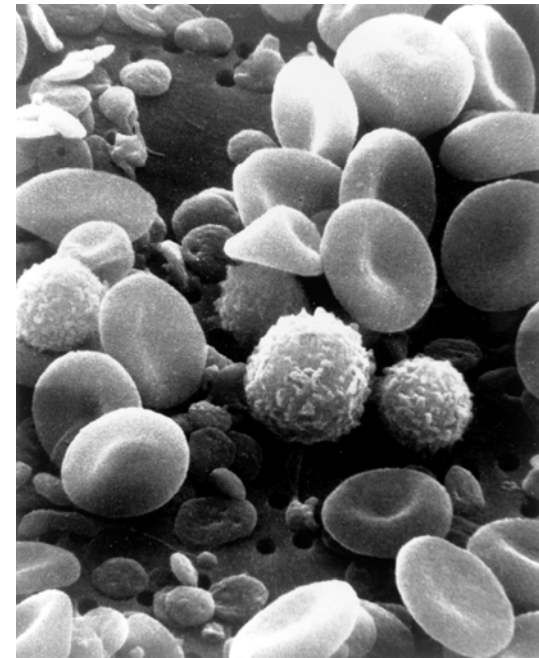
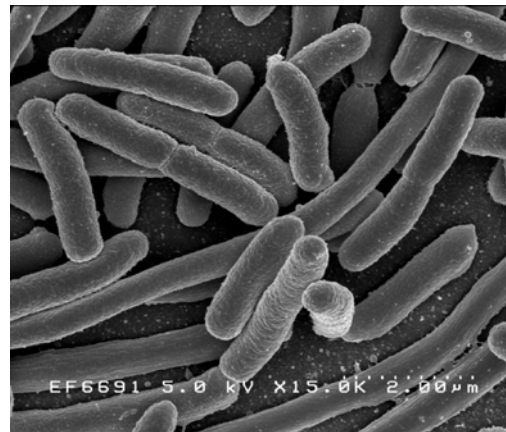
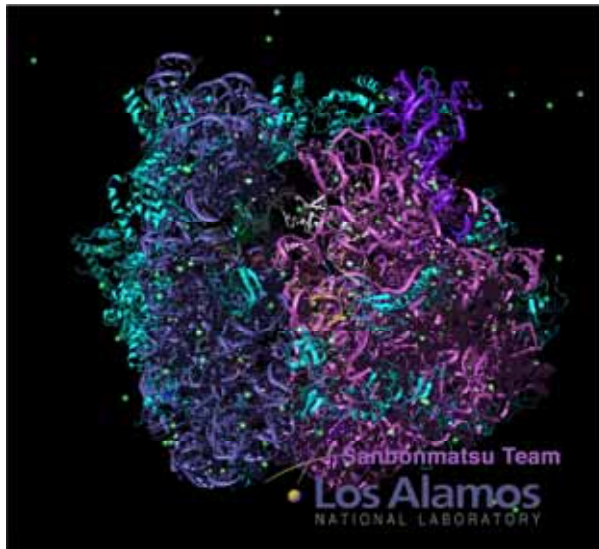


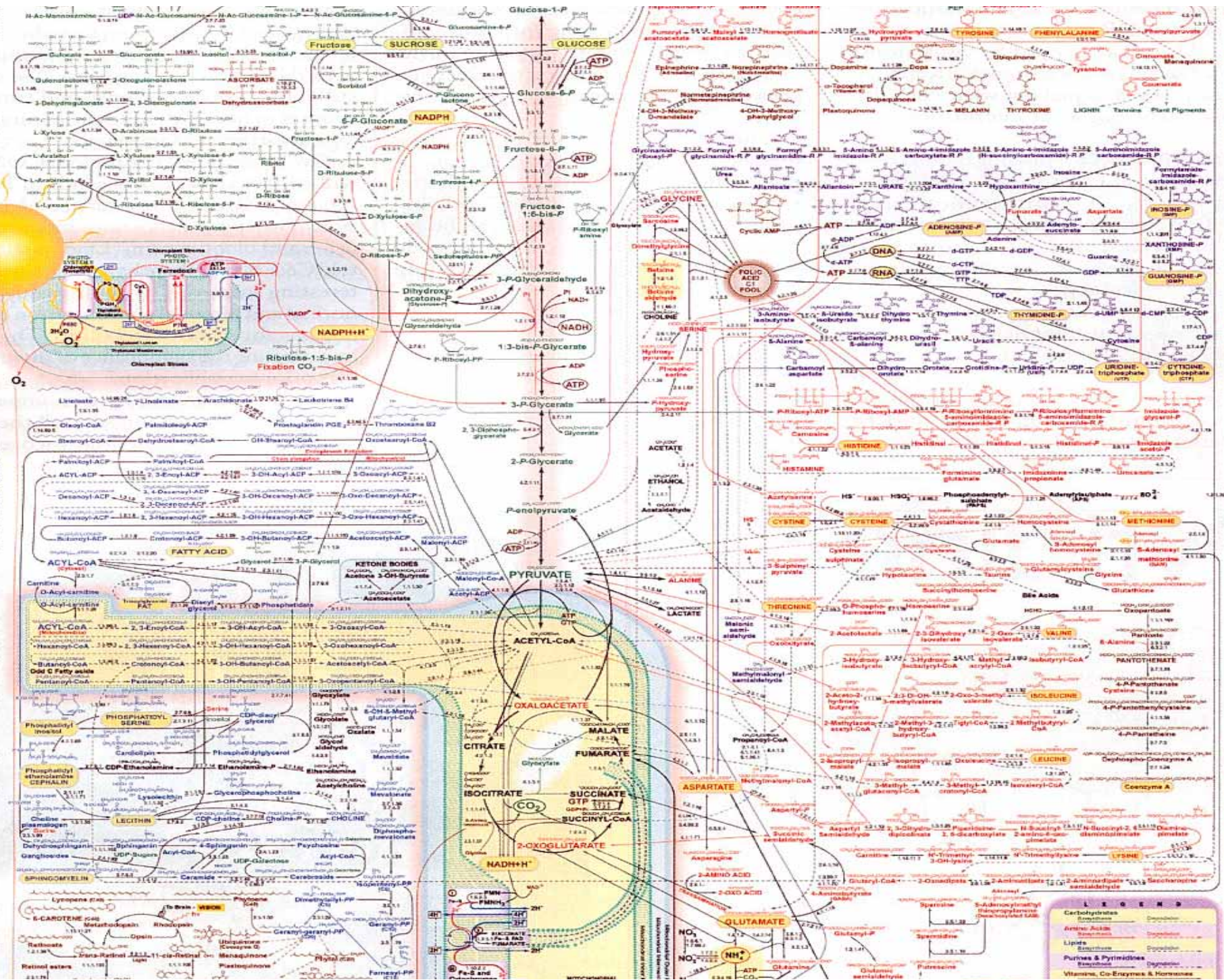


Scale is important !



Biosystems
 $10E-10 \rightarrow 10E7m$

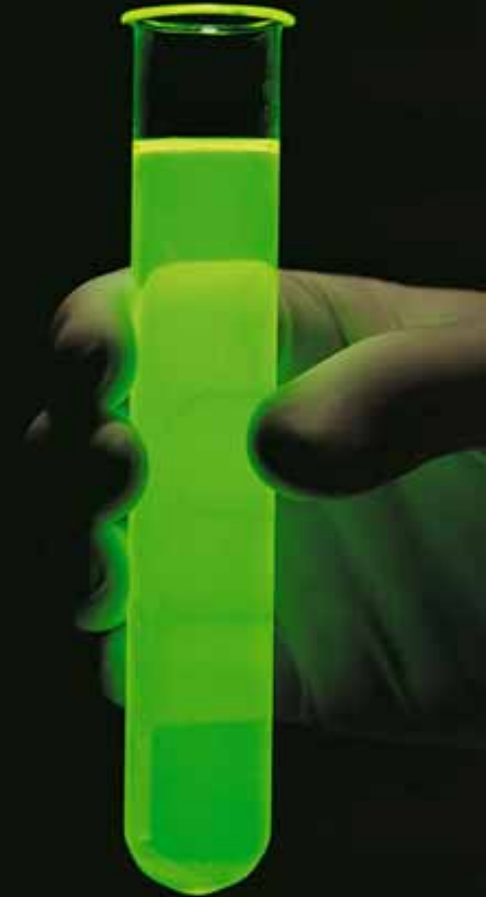
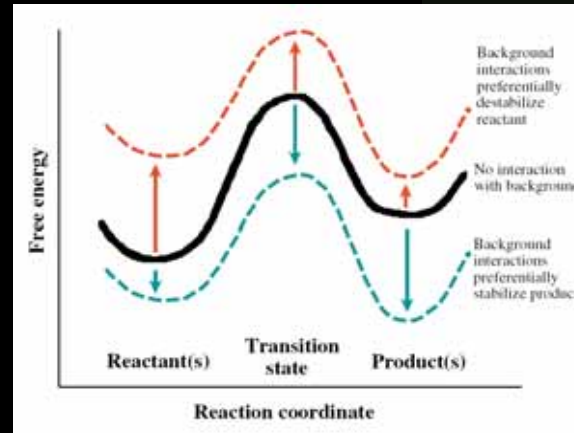
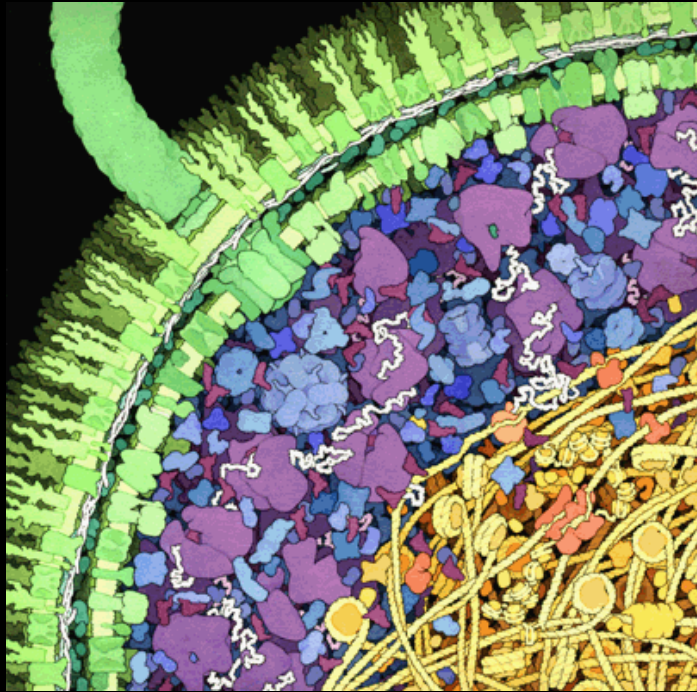




LEGEND

- Carbohydrates
- Amino Acids
- Lipids
- Purines & Pyrimidines
- Vitamins, Co-Enzymes & Hormones

Macromolecular crowding



How can biochemical reactions within cells differ from those in test tubes?

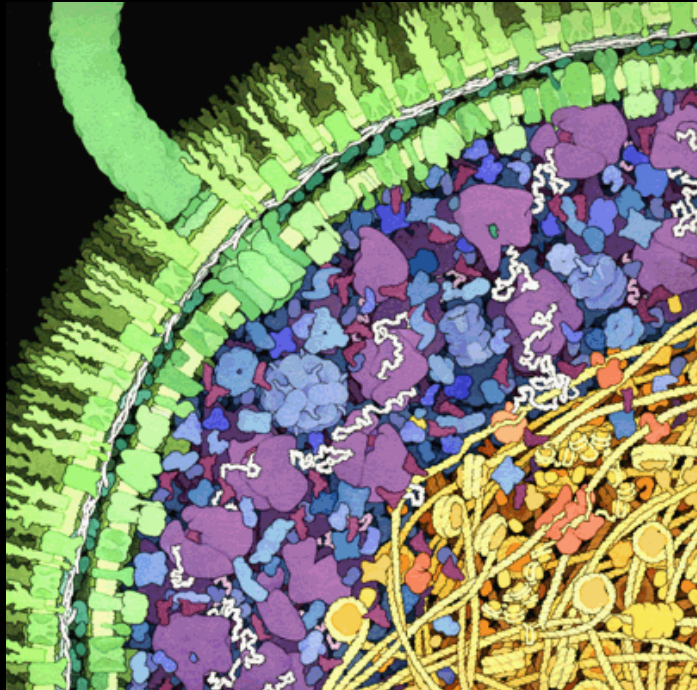
Allen P. Minton

Section on Physical Biochemistry, Laboratory of Biochemical Pharmacology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD, USA
e-mail: minton@helix.nih.gov

Accepted 23 May 2006
Journal of Cell Science 119, 2863-2869 Published by The Company of Biologists 2006
doi:10.1242/jcs.03063

Jülich Plant Phenotyping Centre

Macromolecular crowding



How can biochemical reactions within cells differ from those in test tubes?

Allen P. Minton

Section on Physical Biochemistry, Laboratory of Biochemical Pharmacology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD, USA
e-mail: minton@helix.nih.gov

Accepted 23 May 2006
Journal of Cell Science 119, 2863-2869 Published by The Company of Biologists 2006
doi:10.1242/jcs.03063

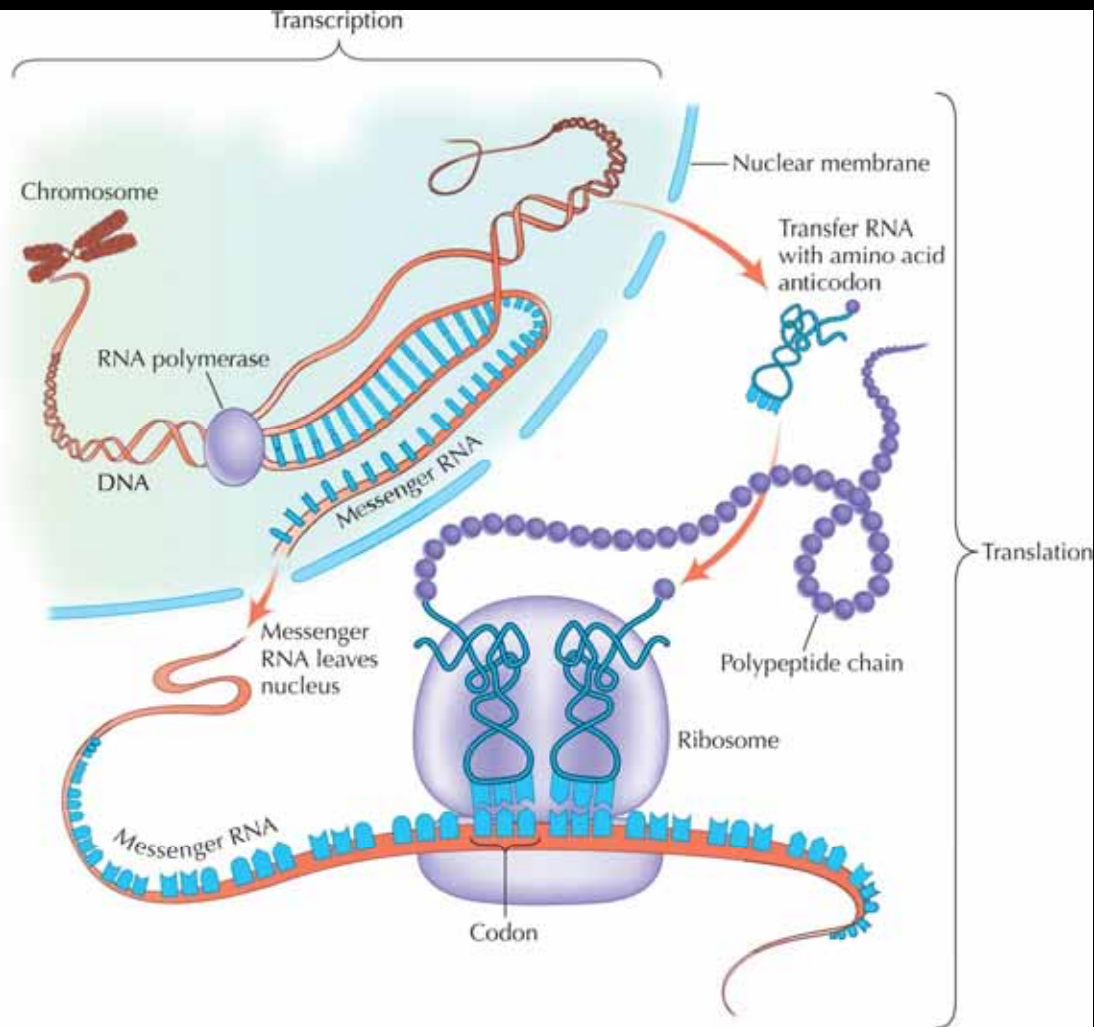
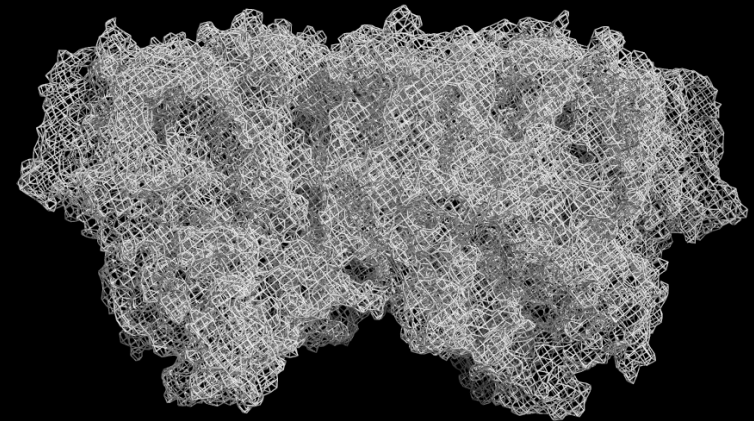
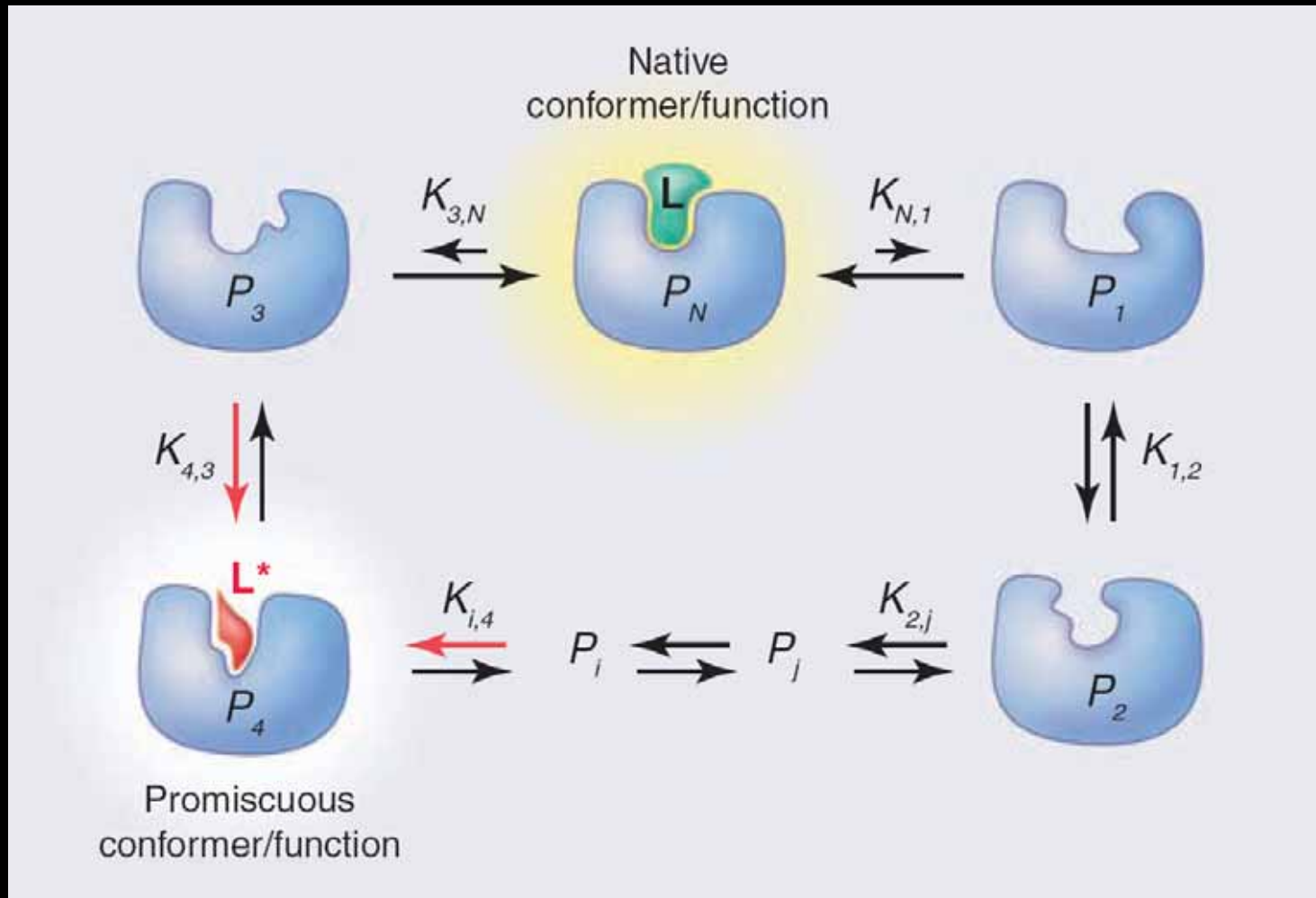


FIGURE 2.27. General mechanism of eukaryotic protein synthesis. The major steps include transcription of the DNA gene sequence into the messenger RNA template in the nucleus of the cell, translation of the DNA codons of that gene into amino acids, and their assembly into polypeptides in the cytoplasm. Important mediators of this process include transfer RNAs, splicing elements, and ribosomes.





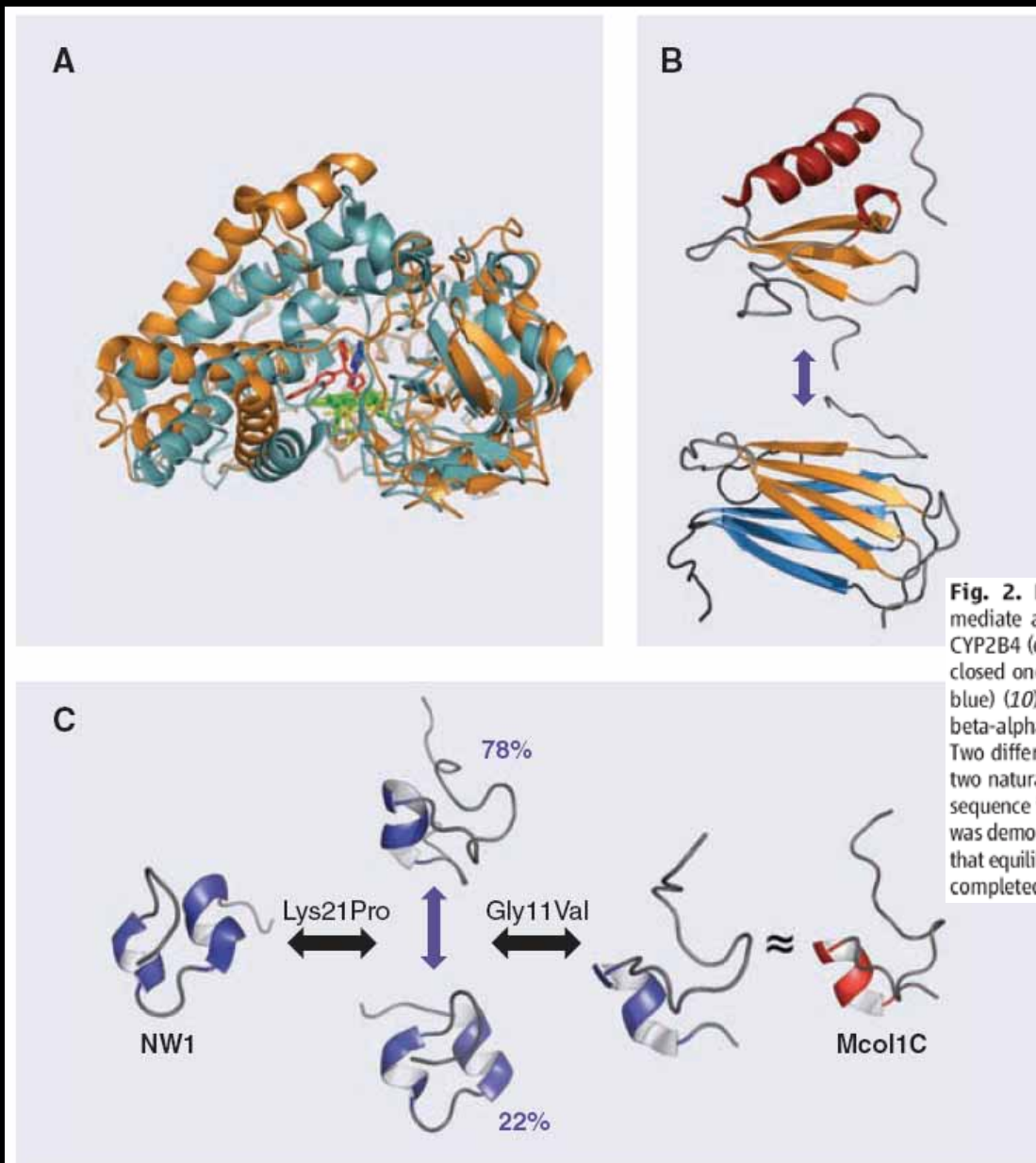
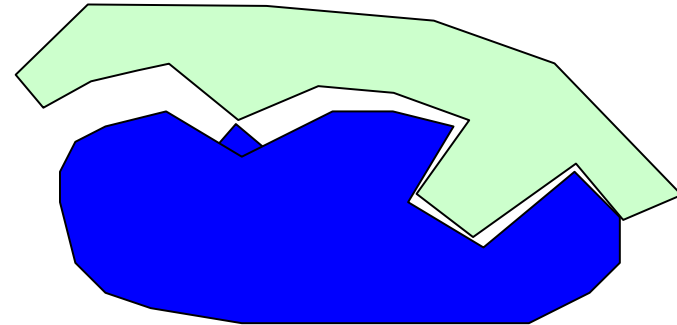
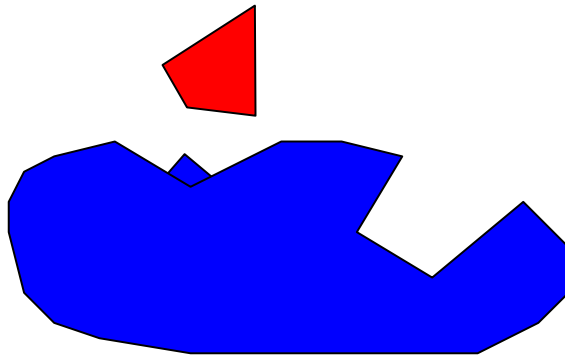
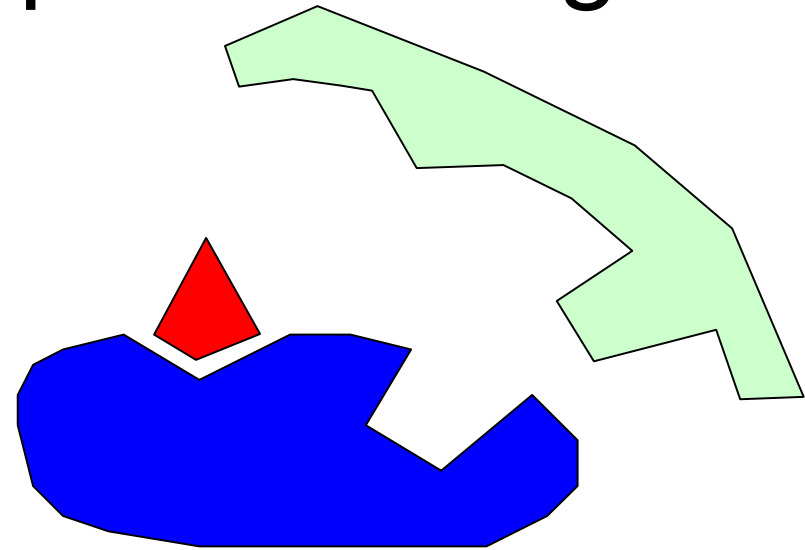
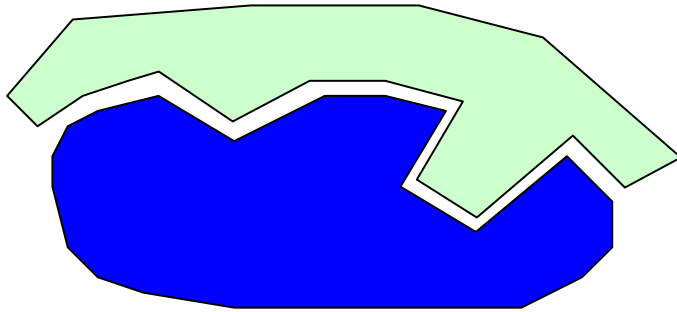
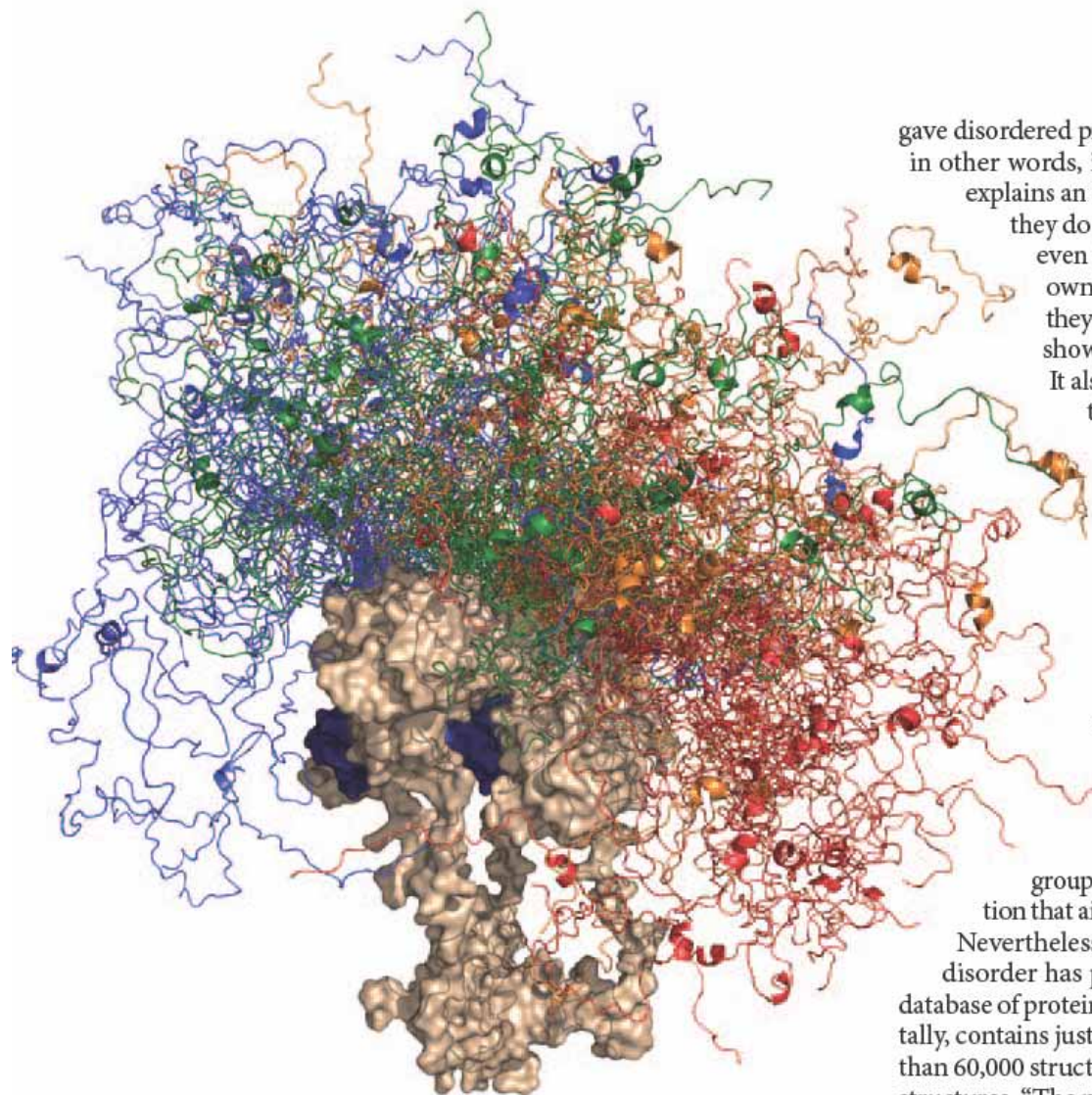


Fig. 2. Examples of conformational diversity. **(A)** Local conformational changes mediate an enzyme's broad substrate specificity. The open conformation of P450-CYP2B4 (orange) occurs with a large substrate (bifonazole, illustrated in red), and the closed one (light blue) occurs with the smaller 4-(4-chlorophenyl) imidazole (darker blue) (10). **(B)** Metamorphic proteins. Lymphotoxin exists in equilibrium between a beta-alpha monomer (top) and an all-beta dimer (bottom) (20). **(C)** Transition folds. Two different topologies (mediated by three different disulfide bridges) are found in two naturally occurring cysteine-rich domains (NW1 and Mcol1C) that show almost no sequence identity beyond the conserved cysteines. Conversion between these topologies was demonstrated via one mutation $\text{Lys}^{21} \rightarrow \text{Pro}^{21}$ (K21P) that afforded an intermediate that equilibrates between the two topologies, and a second mutation $\text{Gly}^{11} \rightarrow \text{Val}^{11}$ (G11V) completed the transition (26).

Conformational proofreading



Tawfik & Alon



Tumour-suppressor protein p53 contains an ordered, globular domain (brown); its disordered segments (colours) help it interact with hundreds of partners.

gave disordered proteins a new meaning. In other words, it explains an aspect of protein function that they do not even tell you about. They are showing us a new way of looking at proteins. It also shows that

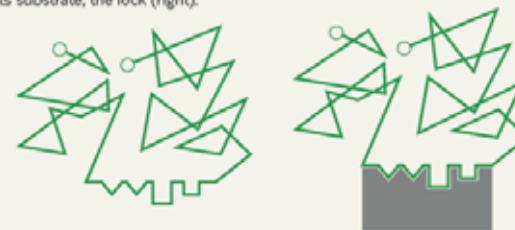
group function. Nevertheless, disorder has provided a new database of proteins. The database, which contains just over 60,000 structures. "The main reason is because it has taken the past few years, and disordered proteins

ORDERS OF DISORDER

A central tenet in molecular biology states that the function of a protein depends on its fully folded three-dimensional structure. In the new view, protein segments can function when transiently or durably disordered.

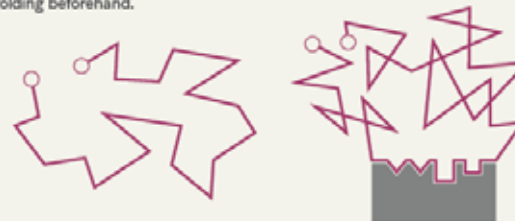
LOCK AND KEY

In the conventional view, an enzyme folds up immediately into a unique and stable 3D shape, the key (left). Its shape perfectly matches and allows it to bind its substrate, the lock (right).



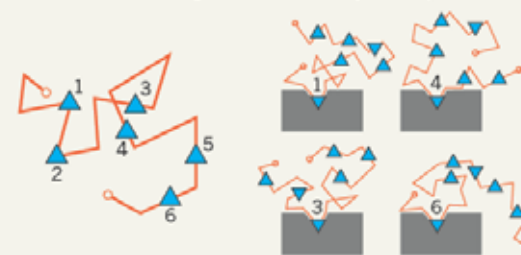
FOLD AS YOU BIND

A disordered part of the gene-regulatory protein CREB (left) uses the lock to mould itself into the shape of the key when the two meet (right), rather than folding beforehand.

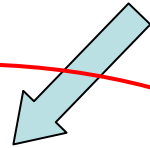


SHAPE SHIFTING

The signalling protein Sic1 remains disordered in its bound state, and each of six phosphate groups occupies the binding site in turn. The protein is a mix of different conformations shifting around in constant dynamic equilibrium.

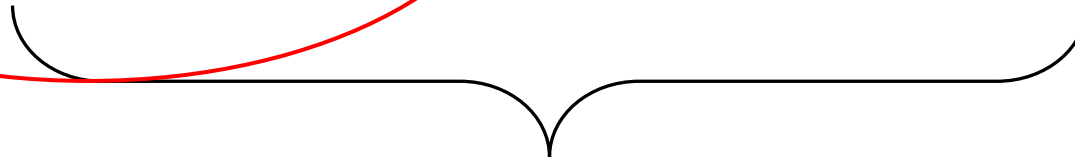


Protein Engineering

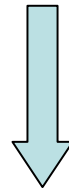


Rational Design
(Directed Mutagenesis)

Random Mutagenesis/
Directed Evolution



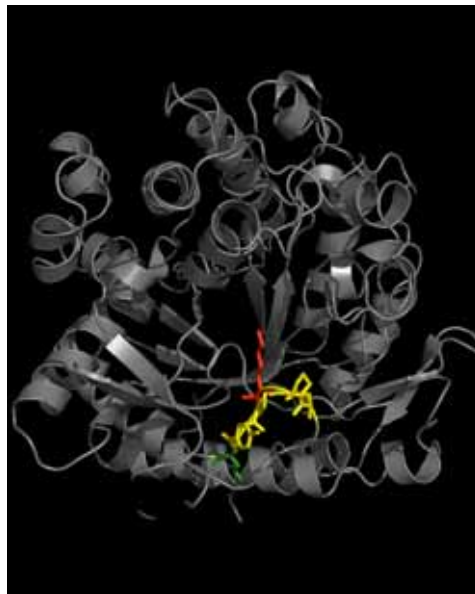
Improved Protein



Protein Engineering



Rational Design (Directed Mutagenesis)



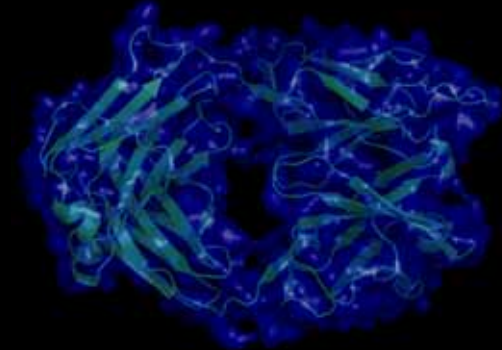
Google: AdvGentech5.ppt

Bioinformatics

```

DBID[CHAIN]SVTLTQSPSSLSASLGDTITTCASQNIHVWLSWYQKPKGHIPLKLIYA
A0A5D8 ROUSE IQMHSQSPSSLSASLGDTITTCASQNIHVWLSWYQKPKGHIPLKLIYA
Q58EU4 ROUSE IVMTQSHKHSTSLGDRVSTTCASQDVTAVAWYQKPGSPKVLIIYA
Q77S96 ROUSE IGMTQSPSSWZASLGDRVSTTCASQDITSLWYQKPGSPKVLIIYA
A0A5E3 ROUSE IVLTQSPALMSASLGDRVITTCASSVSYLHWYQKPGSPKLIYA
A0A5E6 ROUSE IVLTQSPALMSASLGDRATISCKASQSDVYHWYQKPGSPKLIYA
Q66J57 ROUSE IVLTQSPALMSASLGDRATISCKASQSVSYLHWYQKPGSPKLIYA
Q52L54 ROUSE IVMSQSPSSLSASLGDRVITTCASQDITSLWYQKPGSPKLIYA
A0A5E0 ROUSE IVLTQSPALMSASPGDRVITTCASSSVSYLHWYQKPGSPKLIYA
Q569Y8 ROUSE IVLTQSPALMAASPGERVITTCSSSSIDSILHWYQKSGTSPKWIYV
Q52L95 ROUSE IGMTQSPSSWYVSLGEAVITTCASQDITSLWYQKPGSPKLIYA
Q56EV6 ROUSE IVLTQSPALMSASPGDRVITTCASSSVSYLHWYQKSGTSPKWIYV
A0N7J3 RHUR ILMITQPLSLPVSLDQASISCKSSQSLNTLYHWLQKPGSPKLIYS
Q52FY8 ROUSE IVLTQSPALMSASPGDRVITTCASSSVSYLHWYQKPGSPKLIYA
Q3KQK1 ROUSE IVLTQSPALMAASPGERVITTCSSSSIDSILHWYQKSGTSPKWIYV
A0A5D7 ROUSE IVMTQPLSLPVSLDQASISCKSSQSLNTLYHWLQKPGSPKLIYS
Q65ZC0 ROUSE IVMTQPLSLPVSLDQASISCKSSQSLNTLYHWLQKPGSPKLIYS
Q52K04 ROUSE IVMTQAAPSVPTPDESVSISCKSSQSLNTLYHWLQKPGSPKLIYS
Q58EU8 ROUSE IVMTQPLSLPVSLDQASISCKSSQSLNTLYHWLQKPGSPKLIYS
A0A5D9 ROUSE IVMTQPLSLPVSLDQASISCKSSQSLNTLYHWLQKPGSPKLIYS
Q4KRS6 RAT IGMTQSPSSLSASLGDTITTCASQNIHVWLSWYQKPKGHIPLKLIYA
    
```

DNA and AA sequence information:
composition, similarity,
physicochemical properties,
motifs, experimental design for directed
or directed evolution procedure



3D structure derived information:
structure motifs, similarity,
surfaces, canals and cavities,
check and evaluation of aminoacids
possible involved in misfolding



Literature search: current knowledge in this area,
data mining, identification of possible problems,
look for collaborations material etc.

HotSpot Wizard 1.7

HotSpot Wizard

loschmidt.chemi.muni.cz/hotspotwizard/

Genesis 1:1 ASV - Yo... Google Bookmark Manager SquirrelMail - Přihlásit Social Media News an... Wikipedia, the free e... Seznam Slovník - Vícej... Other bookmarks

test2 (1hxj)

[Back to results](#)

[Submit new job](#)

[Download results](#) in one package

- Download individual files
- Links to databases and tools

+ Color Coding

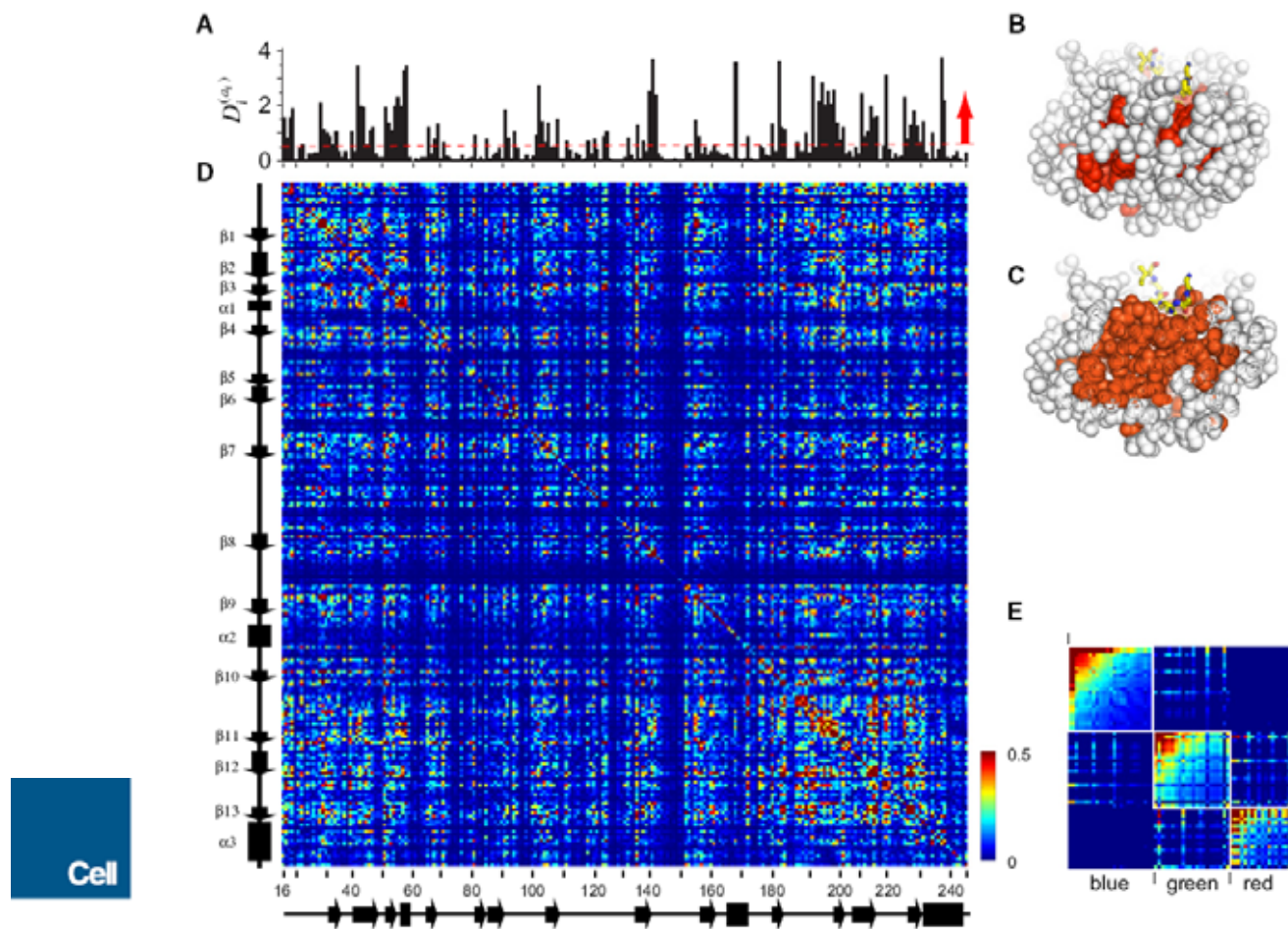
+ ⚡ Controls

Sequence

1hxj

Chain A:

10	20	30	40	50
QMLSPSEIPQRDWFPDFTFGAATSAYQIEGAWNEDGKGESNWDHFCNH				
60	70	80	90	100
PERILDGNSNDIGANSYHMYKTDVRLLEKMGMDAYRFSISWPRILPKGTK				
110	120	130	140	150
EGGINPDGIKYRNLIINLLENGIEPYVTIFHWDPQALEEKYGGFLDKS				
160	170	180	190	200
HKSIVEDYTYFAKVCDFNFGDKVKNULTFNEPQTFTFSYGTGVFAAGRC				
210	220	230	240	250
SPGLDCAYPTGNSLVEPYTAGHNILLAHAEAVDLYNKHYKRDDTRIGLAF				
260	270	280	290	300
DVMGPNRYGTSFLDKQLEERSNDNLGHELEPNHRCYRESMPSLAPRI				



Protein Sectors: Evolutionary Units of Three-Dimensional Structure

Najeeb Halabi,^{1,4} Olivier Rivoire,^{2,4} Stanislas Leibler,^{2,3} and Rama Ranganathan^{1,*}

¹The Green Center for Systems Biology, and Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX 75390-9050, USA

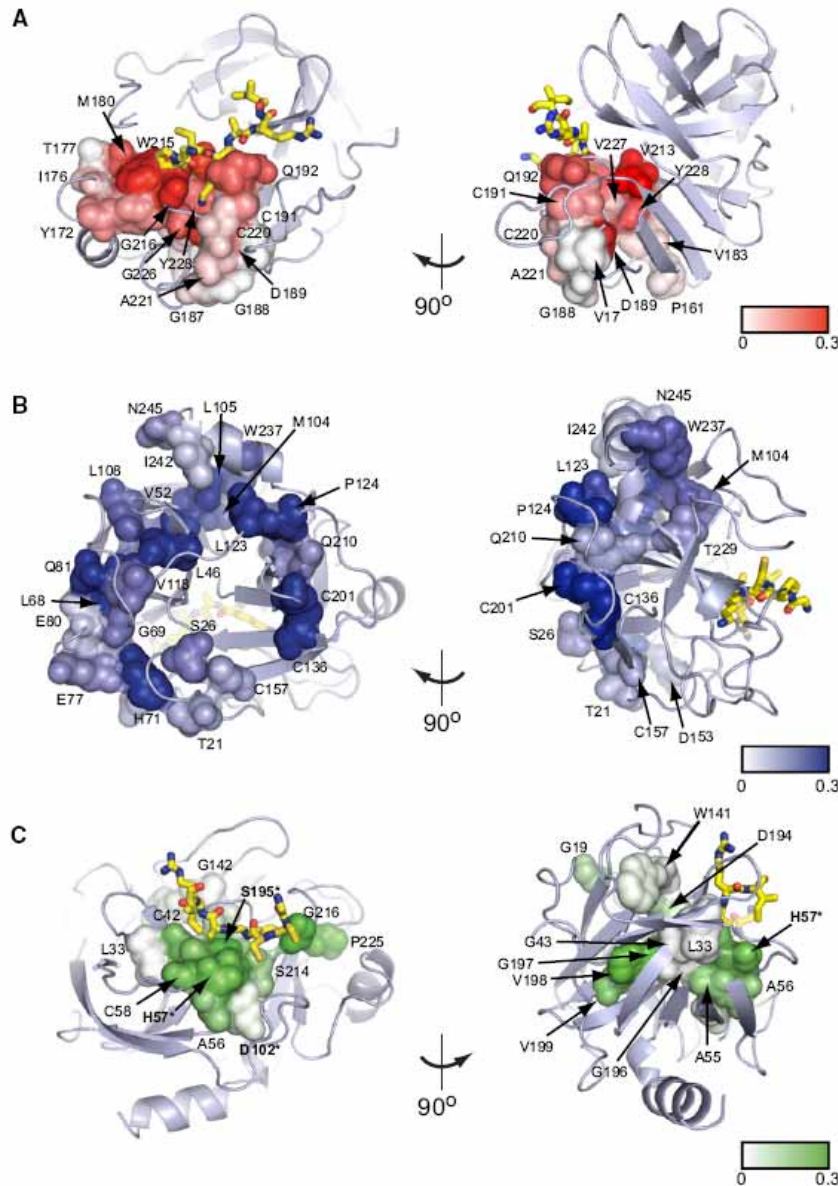
²The Center for Studies in Physics and Biology and Laboratory of Living Matter, Rockefeller University, New York, NY 10065, USA

³The Simons Center for Systems Biology and the School of Natural Sciences, The Institute for Advanced Study, Princeton, NJ 08540, USA

⁴These authors contributed equally to this work

*Correspondence: rama.ranganathan@utsouthwestern.edu

DOI 10.1016/j.cell.2009.07.038

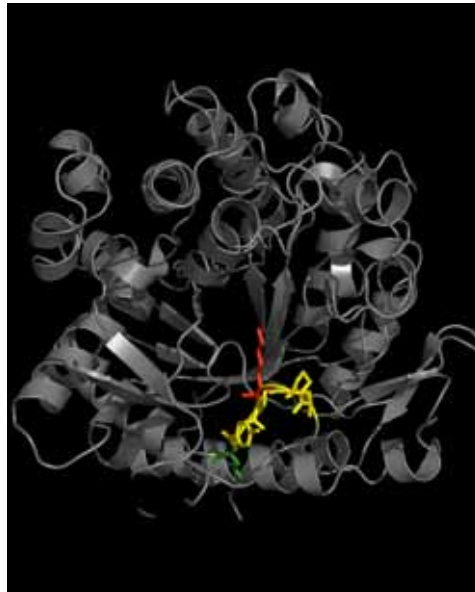


(1) protein domains have a heterogeneous internal organization of amino acid interactions that can comprise multiple functionally distinct subdivisions (the sectors), and

(2) these sectors define a decomposition of proteins that is distinct from the hierarchy of primary, secondary, tertiary, and quaternary structure. We propose that the sectors are features of protein structures that reflect the evolutionary histories of their conserved biological properties

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

Protein Engineering



Google: AdvGentech5.ppt

Protein Engineering

What can be engineered in Proteins ?

-> **Folding (+Structure):**

1. **Thermodynamic Stability**
(Equilibrium between: Native \Leftrightarrow Unfolded state)
2. **Thermal and Environmental Stability** (Temperature, pH, Solvent, Detergents, Salt

Protein Engineering

What can be engineered in Proteins ?

-> Function:

1. **Binding** (Interaction of a protein with its surroundings)

How many points are required to bind a molecule with high affinity?

1. **Catalysis** (a different form of binding – binding the transition state of a chemical reaction)

Increased binding to the transition state \Rightarrow increased catalytic rates !!!

Requires: Knowledge of the Catalytic Mechanism !!!

-> engineer K_{cat} and K_m

Protein Engineering

Factors which contribute to stability:

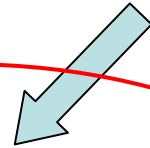
1. Hydrophobicity (hydrophobic core)
1. Electrostatic Interactions:
 - > Salt Bridges
 - > Hydrogen Bonds
 - > Dipole Interactions
1. Disulfide Bridges
1. Metal Binding (Metal chelating site)
1. Reduction of the unfolded state entropy with
X → Pro mutations

Protein Engineering

Design of Thermal and Environmental stability:

1. Stabilization of α -Helix Macroipoles
1. Engineer Structural Motifes (like Helix N-Caps)
1. Introduction of salt bridges
1. Introduction of residues with higher intrinsic properties for their conformational state (e.g. Ala replacement within a α -Helix)
1. Introduction of disulfide bridges
1. Reduction of the unfolded state entropy with
X \rightarrow Pro mutations

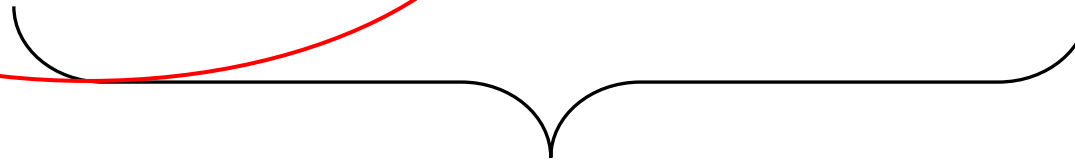
Protein Engineering



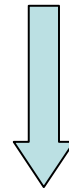
Rational Design
(Directed Mutagenesis)



Random Mutagenesis/
Directed Evolution



Improved Protein



Protein Engineering



		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

Mutagenesis

Mutagenesis -> change in DNA sequence

-> Point mutations or large modifications

Point mutations (directed mutagenesis):

- Substitution: change of one nucleotide (i.e. A-> C)
- Insertion: gaining one additional nucleotide
- Deletion: loss of one nucleotide

Consequences of point mutations within a coding sequence (gene) for the protein

(a) Point mutations and small deletions

Wild-type sequences

Amino acid	N-Phe	Arg	Trp	Ile	Ala	Asn-C
mRNA	5'-UUU	CGA	UGG	AUA	GCC	AAU-3'
DNA	3'-AAA	GCT	ACC	TAT	CGG	TTA-5'
	5'-TTT	CGA	TGG	ATA	GCC	AAT-3'

Silent mutations:

-> change in nucleotide sequence with **no consequences** for protein sequence

Missense

3'-AAT	GCT	ACC	TAT	CGG	TTA-5'
5'-TTA	CGA	TGG	ATA	GCC	AAT-3'
N-Leu	Arg	Trp	Ile	Ala	Asn-C

-> Change of amino acid

Nonsense

3'-AAA	GCT	ATC	TAT	CGG	TTA-5'
5'-TTT	CGA	TAG	ATA	GCC	AAT-3'
N-Phe	Arg	Stop			

-> truncation of protein

Frameshift by addition

3'-AAA	GCT	ACC	ATA	TCG	GTT A-5'
5'-TTT	CGA	TGG	TAT	AGC	CAAT-3'
N-Phe	Arg	Trp	Tyr	Ser	Gln

-> change of c-terminal part of protein

Frameshift by deletion

GCTA					
CGAT					
3'-AAA	↓ CCT	ATC	GGT	TA-5'	
5'-TTT	GGA	TAG	CCA	AT-3'	
N-Phe	Gly	Stop			

-> change of c-terminal part of protein

Codon Usage is different in different species

UUU F 0.57	UCU S 0.11	UAU Y 0.53	UGU C 0.42
UUC F 0.43	UCC S 0.11	UAC Y 0.47	UGC C 0.58
UUA L 0.15	UCA S 0.15	UAA * 0.64	UGA * 0.36
UUG L 0.12	UCG S 0.16	UAG * 0.00	UGG W 1.00
CUU L 0.12	CCU P 0.17	CAU H 0.55	CGU R 0.36
CUC L 0.10	CCC P 0.13	CAC H 0.45	CGC R 0.44
CUA L 0.05	CCA P 0.14	CAA Q 0.30	CGA R 0.07
CUG L 0.46	CCG P 0.55	CAG Q 0.70	CGG R 0.07
AUU I 0.58	ACU T 0.16	AAU N 0.47	AGU S 0.14
AUC I 0.35	ACC T 0.47	AAC N 0.53	AGC S 0.33
AUA I 0.07	ACA T 0.13	AAA K 0.73	AGA R 0.02
AUG M 1.00	ACG T 0.24	AAG K 0.27	AGG R 0.03
GUU V 0.25	GCU A 0.11	GAU D 0.65	GGU G 0.29
GUC V 0.18	GCC A 0.31	GAC D 0.35	GGC G 0.46
GUA V 0.17	GCA A 0.21	GAA E 0.70	GGA G 0.13
GUG V 0.40	GCG A 0.38	GAG E 0.30	GGG G 0.12

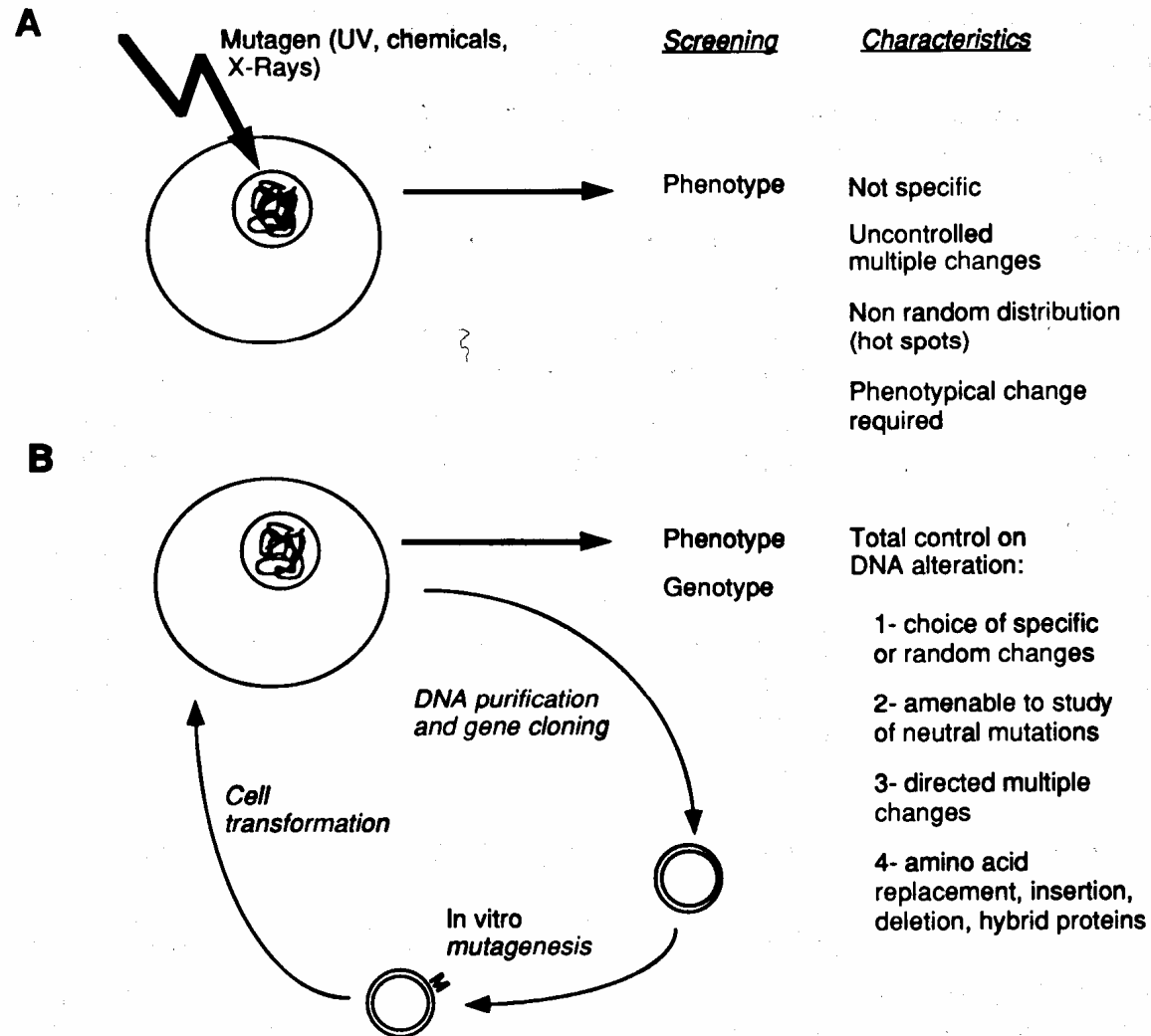
[Codon/a.a./fraction per codon per a.a.]
E. coli K12 data from the Codon Usage Database

UUU F 0.46	UCU S 0.19	UAU Y 0.44	UGU C 0.46
UUC F 0.54	UCC S 0.22	UAC Y 0.56	UGC C 0.54
UUA L 0.08	UCA S 0.15	UAA * 0.30	UGA * 0.47
UUG L 0.13	UCG S 0.05	UAG * 0.24	UGG W 1.00
CUU L 0.13	CCU P 0.29	CAU H 0.42	CGU R 0.08
CUC L 0.20	CCC P 0.32	CAC H 0.58	CGC R 0.18
CUA L 0.07	CCA P 0.28	CAA Q 0.27	CGA R 0.11
CUG L 0.40	CCG P 0.11	CAG Q 0.73	CGG R 0.20
AUU I 0.36	ACU T 0.25	AAU N 0.47	AGU S 0.15
AUC I 0.47	ACC T 0.36	AAC N 0.53	AGC S 0.24
AUA I 0.17	ACA T 0.28	AAA K 0.43	AGA R 0.21
AUG M 1.00	ACG T 0.11	AAG K 0.57	AGG R 0.21
GUU V 0.18	GCU A 0.27	GAU D 0.46	GGU G 0.16
GUC V 0.24	GCC A 0.40	GAC D 0.54	GGC G 0.34
GUA V 0.12	GCA A 0.23	GAA E 0.42	GGA G 0.25
GUG V 0.46	GCG A 0.11	GAG E 0.58	GGG G 0.25

[Codon/a.a./fraction per codon per a.a.]
Homo sapiens data from the Codon Usage Database

Mutagenesis

Comparison of cellular and invitro mutagenesis





The Nobel Prize in Chemistry 1993

"for contributions to the developments of methods within DNA-based chemistry"

"for his invention of the polymerase chain reaction (PCR) method"

"for his fundamental contributions to the establishment of oligonucleotide-based, site-directed mutagenesis and its development for protein studies"



Kary B. Mullis

🏆 1/2 of the prize

USA

La Jolla, CA, USA

b. 1944



Michael Smith

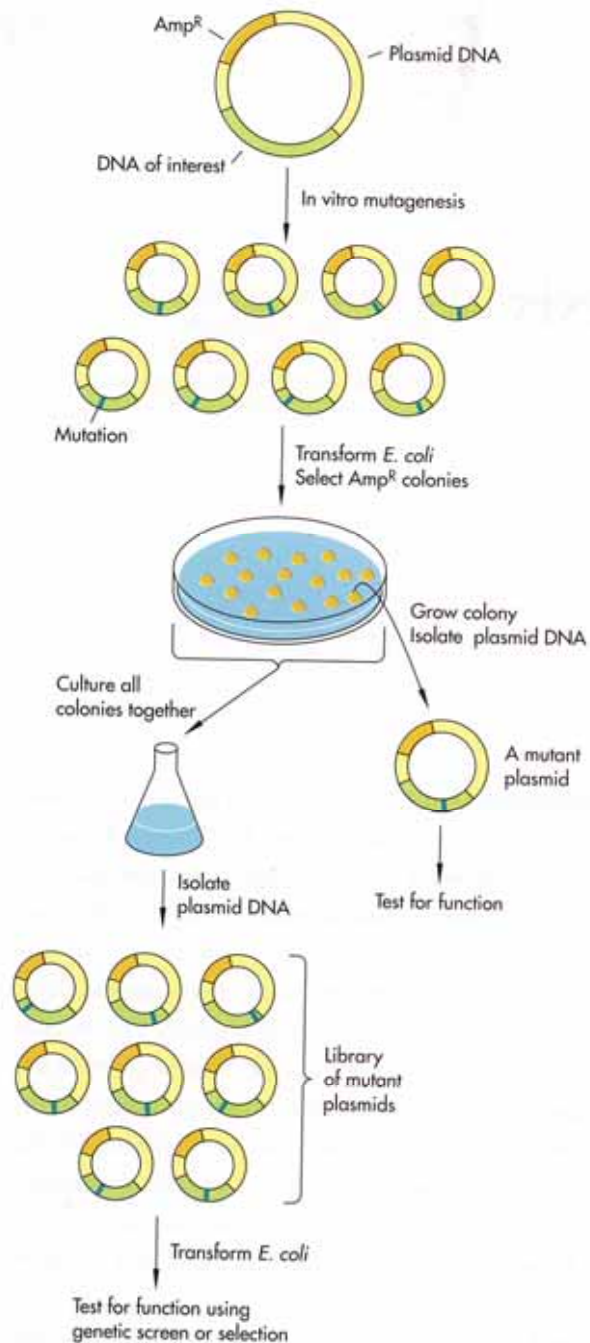
🏆 1/2 of the prize

Canada

University of British
Columbia
Vancouver, Canada

b. 1932
(in Blackpool, United
Kingdom)
d. 2000

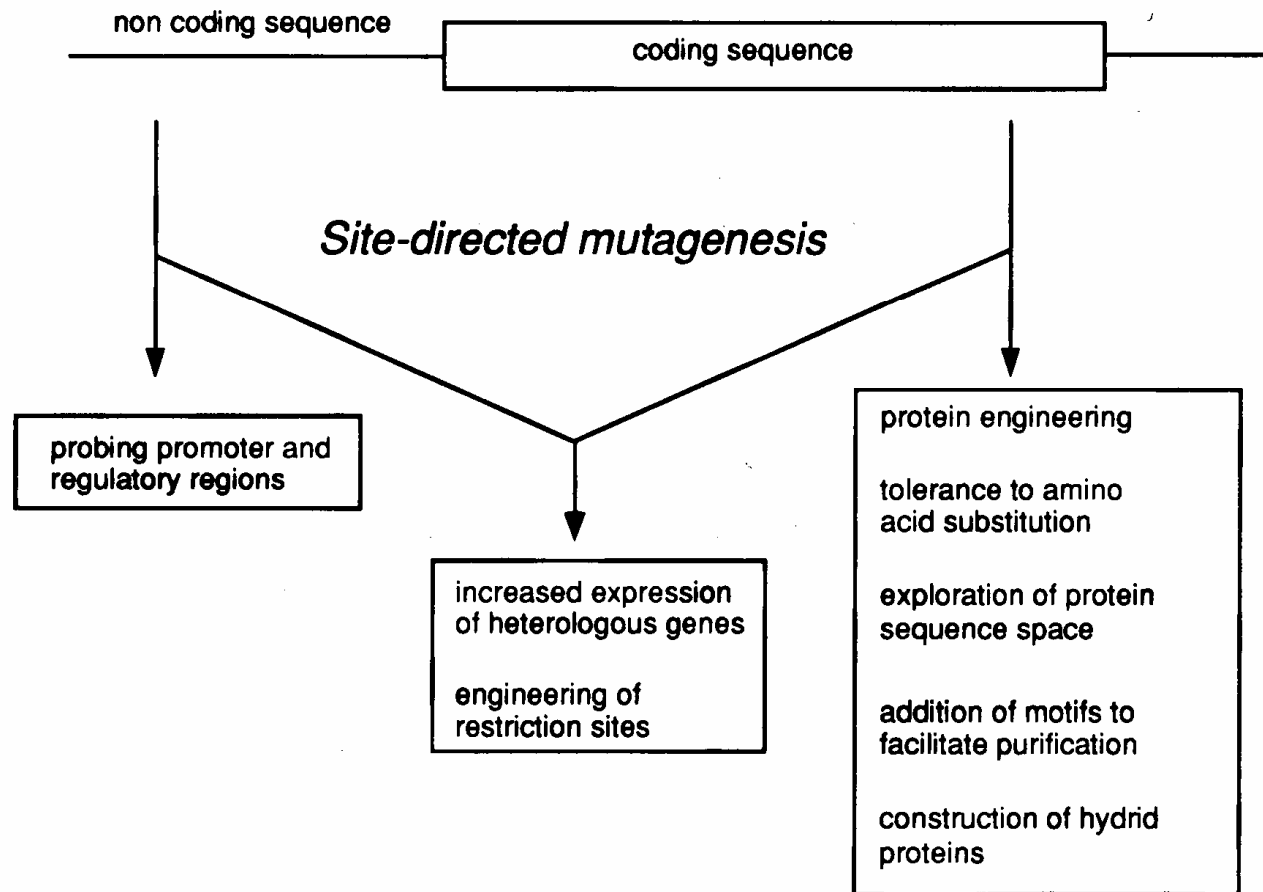
General strategy for directed mutagenesis



Requirements:

- DNA of interest (gene or promoter) must be cloned
- Expression system must be available -> for testing phenotypic change

Applications of directed mutagenesis



Approaches for directed mutagenesis

-> site-directed mutagenesis

-> point mutations in particular known area

result -> library of wild-type and mutated
DNA (site-specific)

not really a library -> just 2 species

Protein Engineering

-> Mutagenesis used for modifying proteins
Replacements on protein level -> mutations on DNA level

Assumption : Natural sequence can be modified to
improve a certain function of protein

This implies:

- Protein is NOT at an optimum for that function
- Sequence changes without disruption of the structure
- (otherwise it would not fold)
- New sequence is not TOO different from the native sequence (otherwise loss in function of protein)

consequence -> introduce point mutations

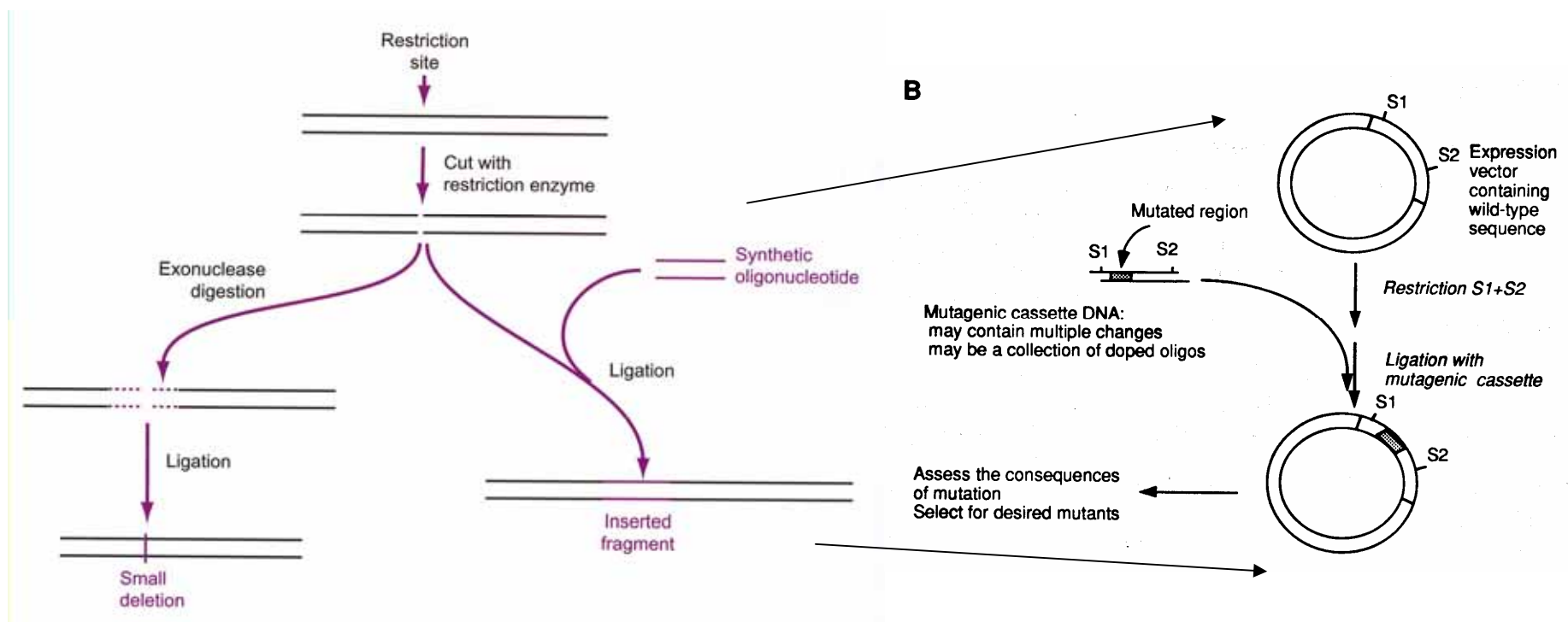
Rational Protein Design

⇒ Site -directed mutagenesis

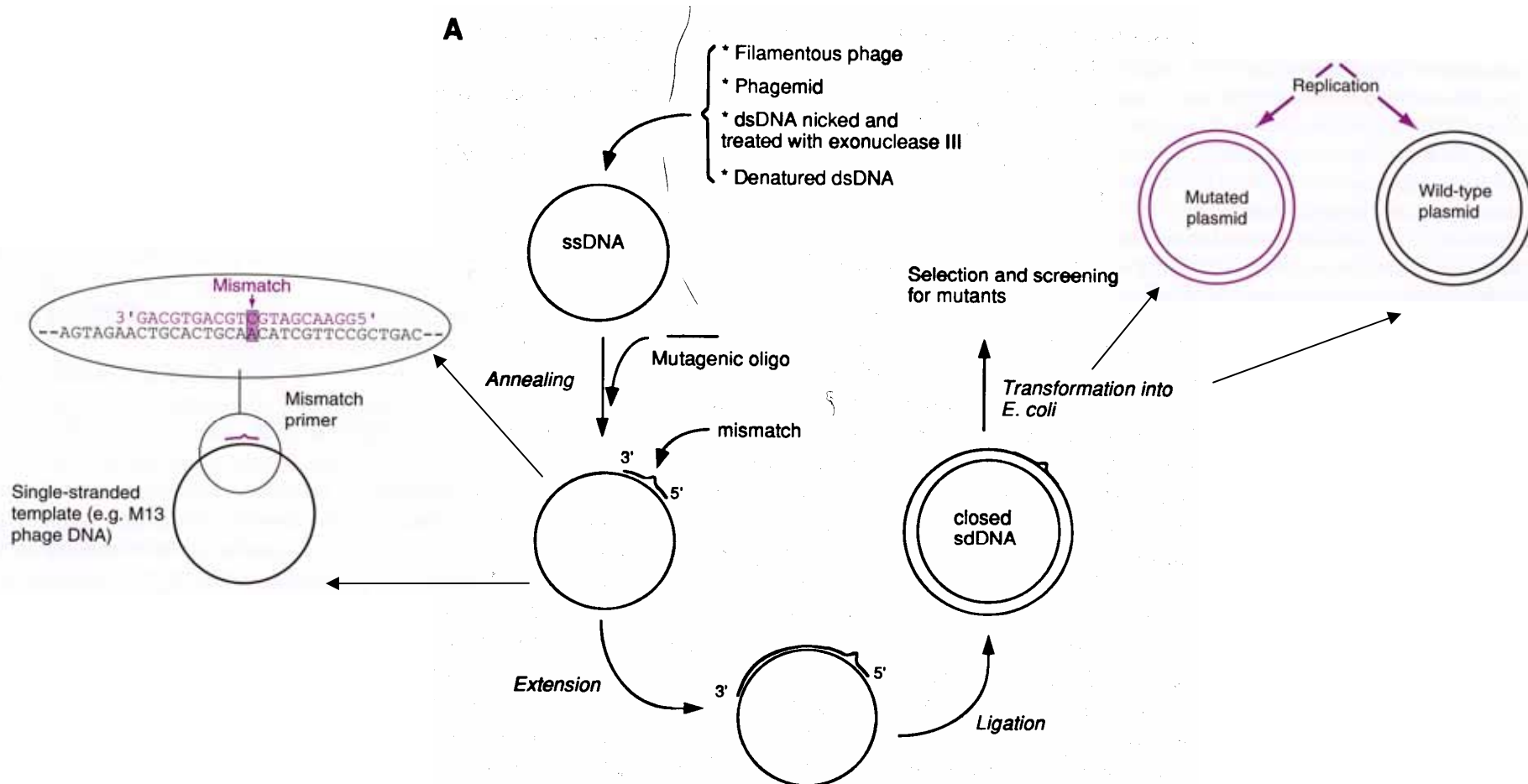
Requirements:

- > Knowledge of sequence and preferable Structure
(active site,....)
- > Understanding of mechanism
(knowledge about structure – function relationship)
- > Identification of cofactors.....

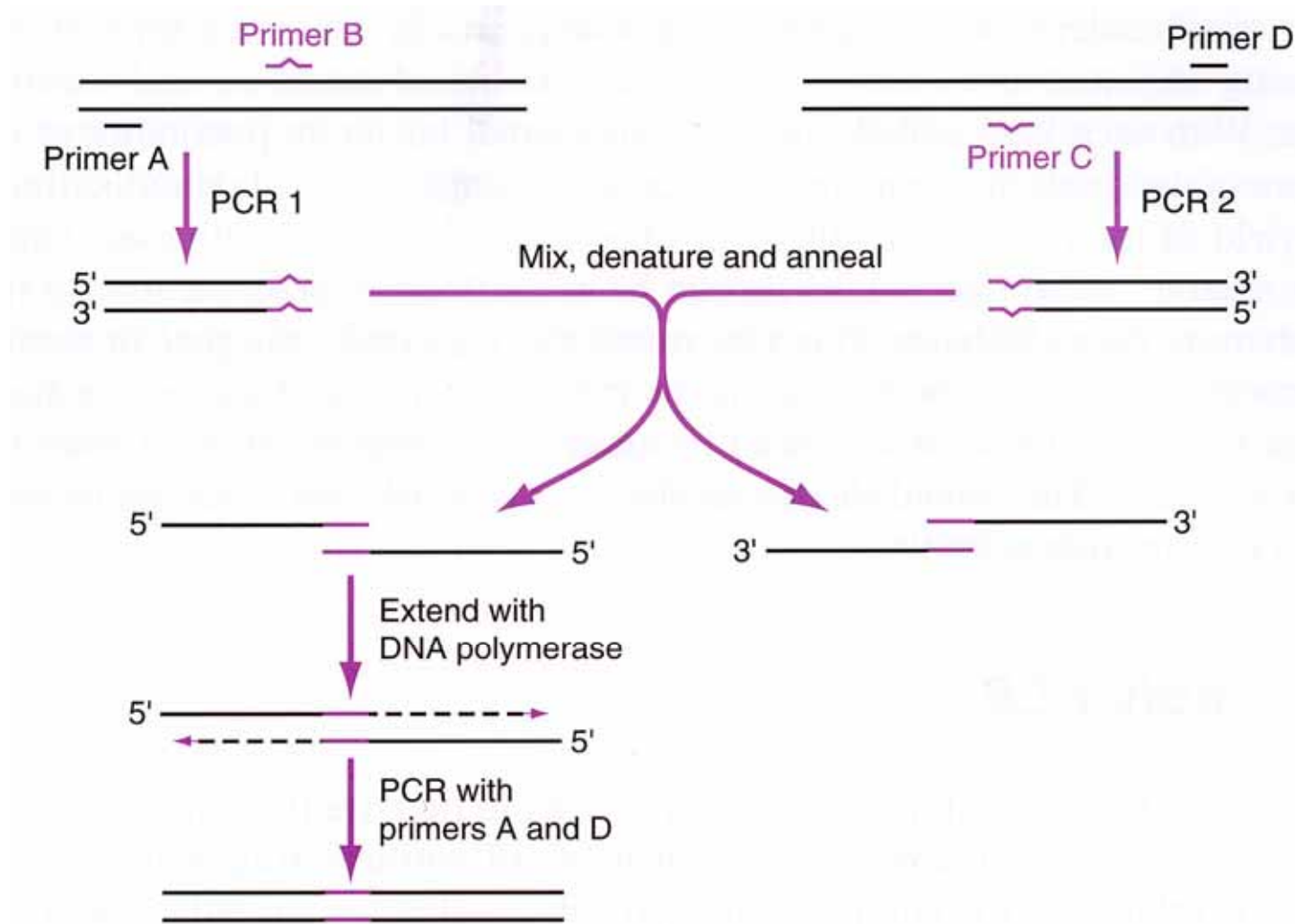
Site-directed mutagenesis methods



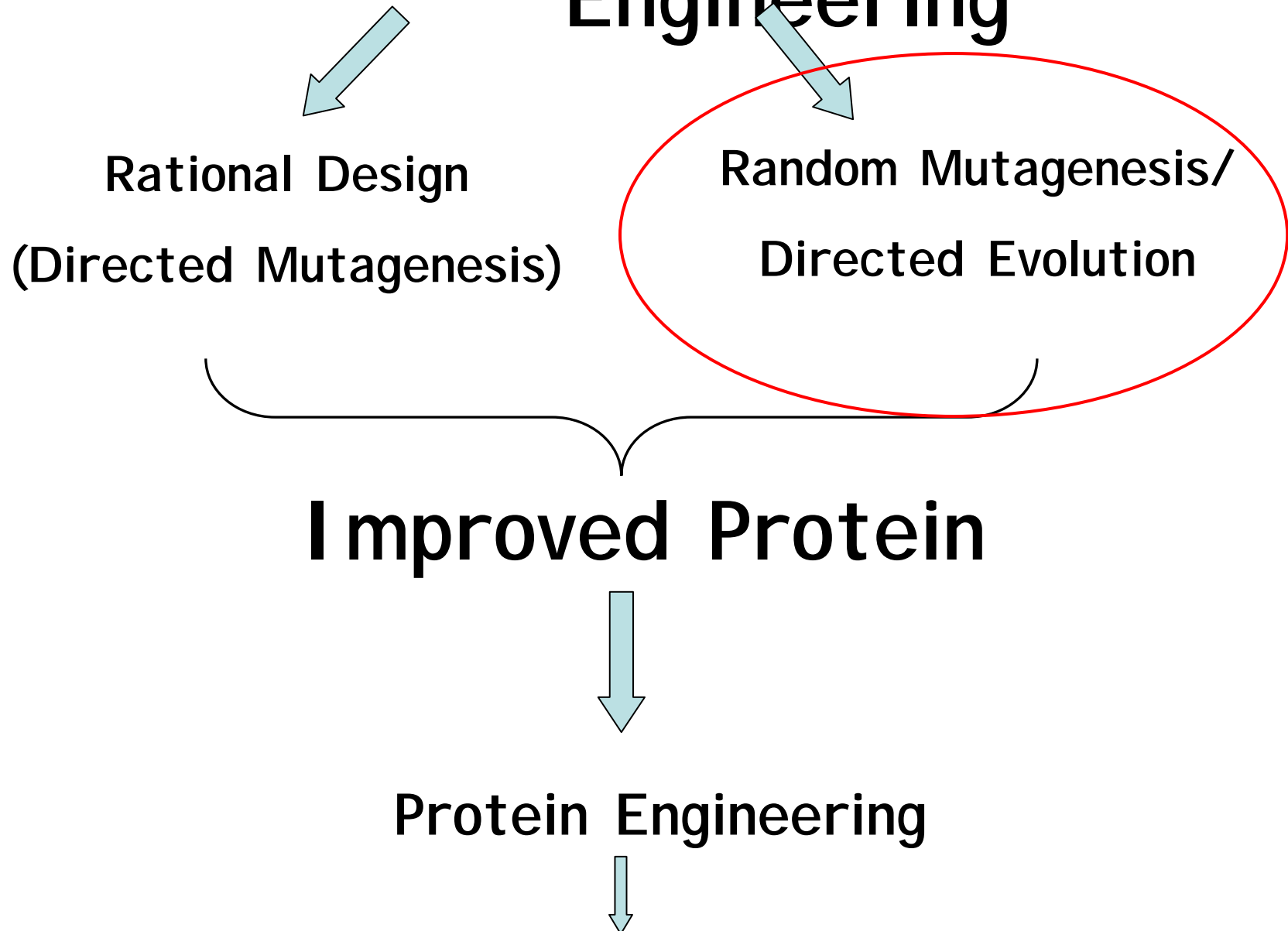
Site-directed mutagenesis methods – Oligonucleotide - directed method



Site-directed mutagenesis methods – PCR based



Protein Engineering



The concept of laboratory-directed protein evolution is not new.

Systematic approaches to directed evolution of proteins have been documented since the 1970s

One early example is the evolution of the EbgA protein from *Escherichia coli*, an enzyme having almost no -galactosidase activity. Through intensive selection of a LacZ deletion strain of *E. coli* for growth on lactose as a sole carbon source, the wild-type EbgA was “evolved” as a -galactosidase sufficient to replace the *lacZ* gene function (Campbell, J.1973).

Directed Evolution – Random mutagenesis

-> based on the process of natural evolution

- NO structural information required
- NO understanding of the mechanism required

General Procedure:

1. Generation of genetic diversity
⇒ Random mutagenesis
2. Identification of successful variants
⇒ Screening and selection

Protein Engineering

Directed Evolution

Successful directed evolution has four requirements:

- (i) the desired function should be physically feasible,
- (ii) the function should be biologically or evolutionary feasible, i.e., a mutational pathway must exist to get from an original protein to tailored protein through ever-improving variants,
- (iii) it should be possible to make libraries of mutants complex enough to contain rare beneficial mutations and
- (iv) a rapid screen or selection reflecting the desired function should be available (Arnold, 1998)

Approaches for directed random mutagenesis

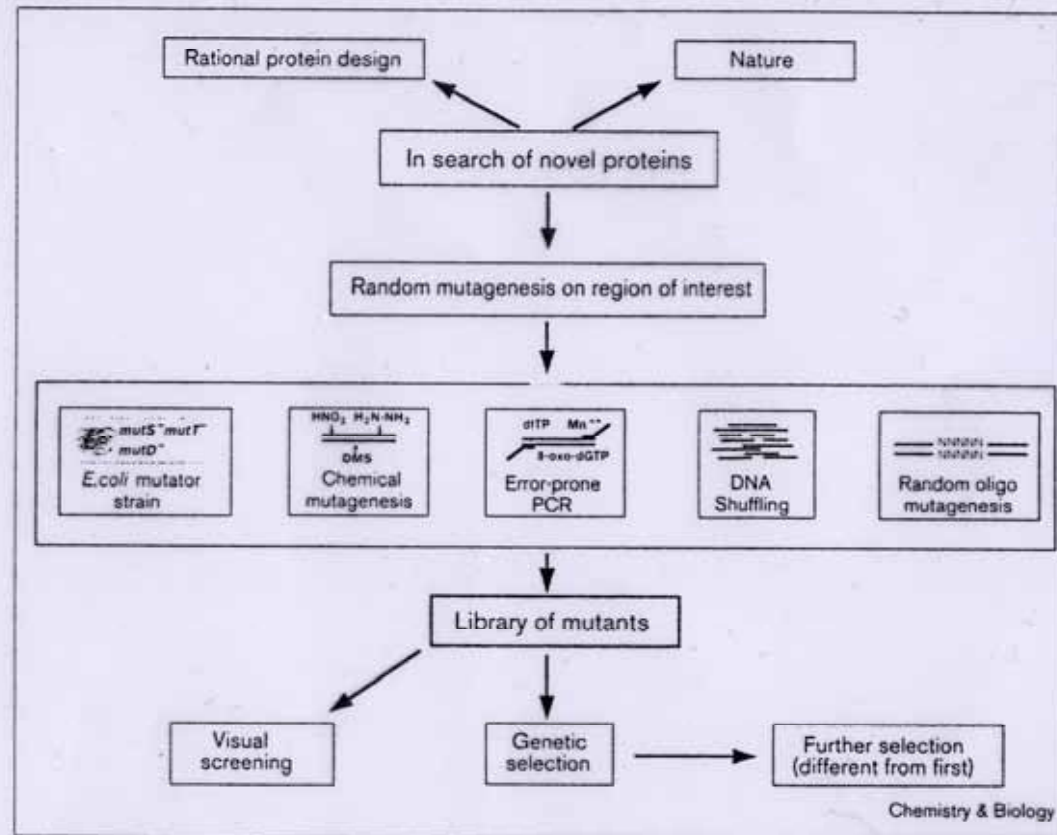
-> random mutagenesis

-> point mutations in all areas within DNA of interest

result -> library of wild-type and mutated DNA (random)
a real library -> many variants -> screening !!!

if methods efficient -> mostly mutated DNA

The creation of random libraries by applied molecular evolution. A gene target can be randomized by one of several methods, followed by ligation of the library into a vector backbone and transformation into an appropriate strain of *Escherichia coli* for selection/screening.



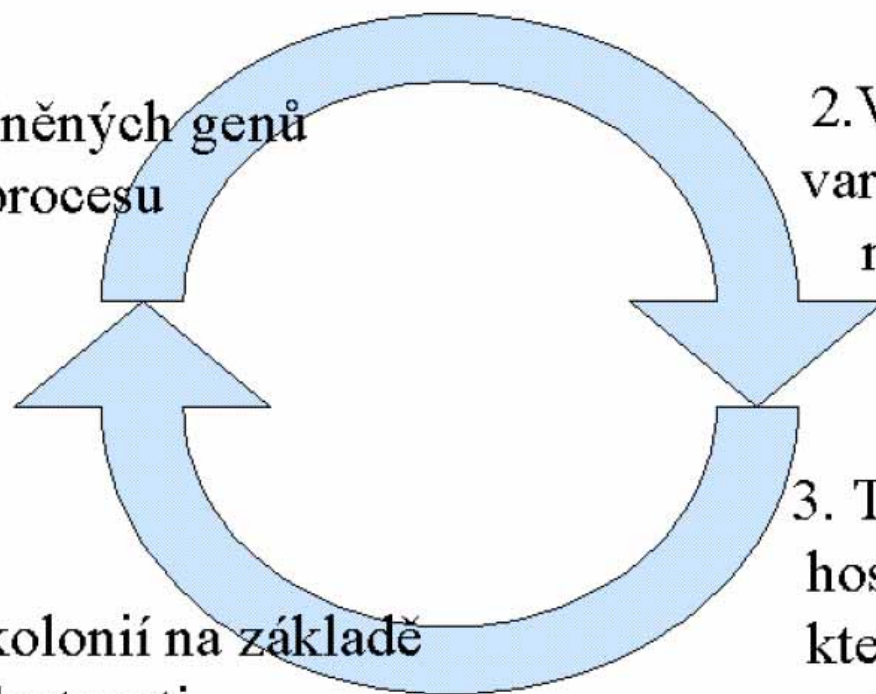
1. Výběr genu zájmu

2. Vytvoření knihovny
variant pomocí mutace
nebo rekombinace

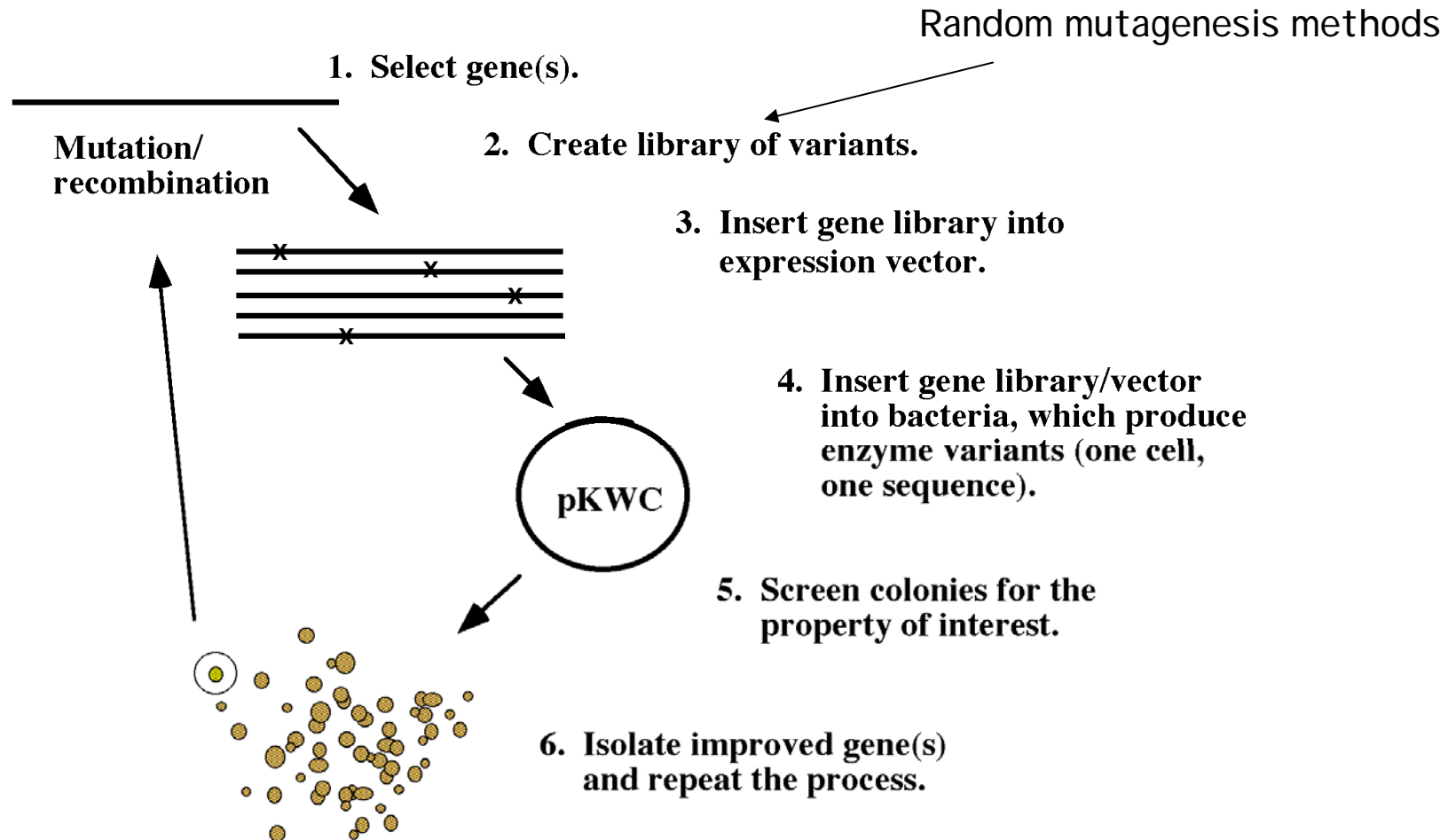
3. Transformace
hostitelského kmene,
který bude produkovat
varianty proteinů

4. Skríníng kolonií na základě
hledané vlastnosti

5. Izolace změněných genů
opakování procesu



General Directed Evolution Procedure



Limitation of Directed Evolution

1. Evolutionary path must exist - > to be successful

1. Screening method must be available

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

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

Typical Directed Evolution Experiment

Successful experiments involve generally
less than 6 steps (cycles)!!!

Why?

1. Sequences with improved properties are rather close to the parental sequence -> along a evolutionary path
2. Capacity of our present methods to generate novel functional sequences is rather limited -> requires huge libraries

⇒ Point Mutations !!!

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
 - > In vivo Recombination (Yeast)
 - > Random priming recombination, Staggered extension process (StEP)
 - > ITCY

Evolutionary Methods

Type of mutation – Fitness of mutants

Type of mutations:

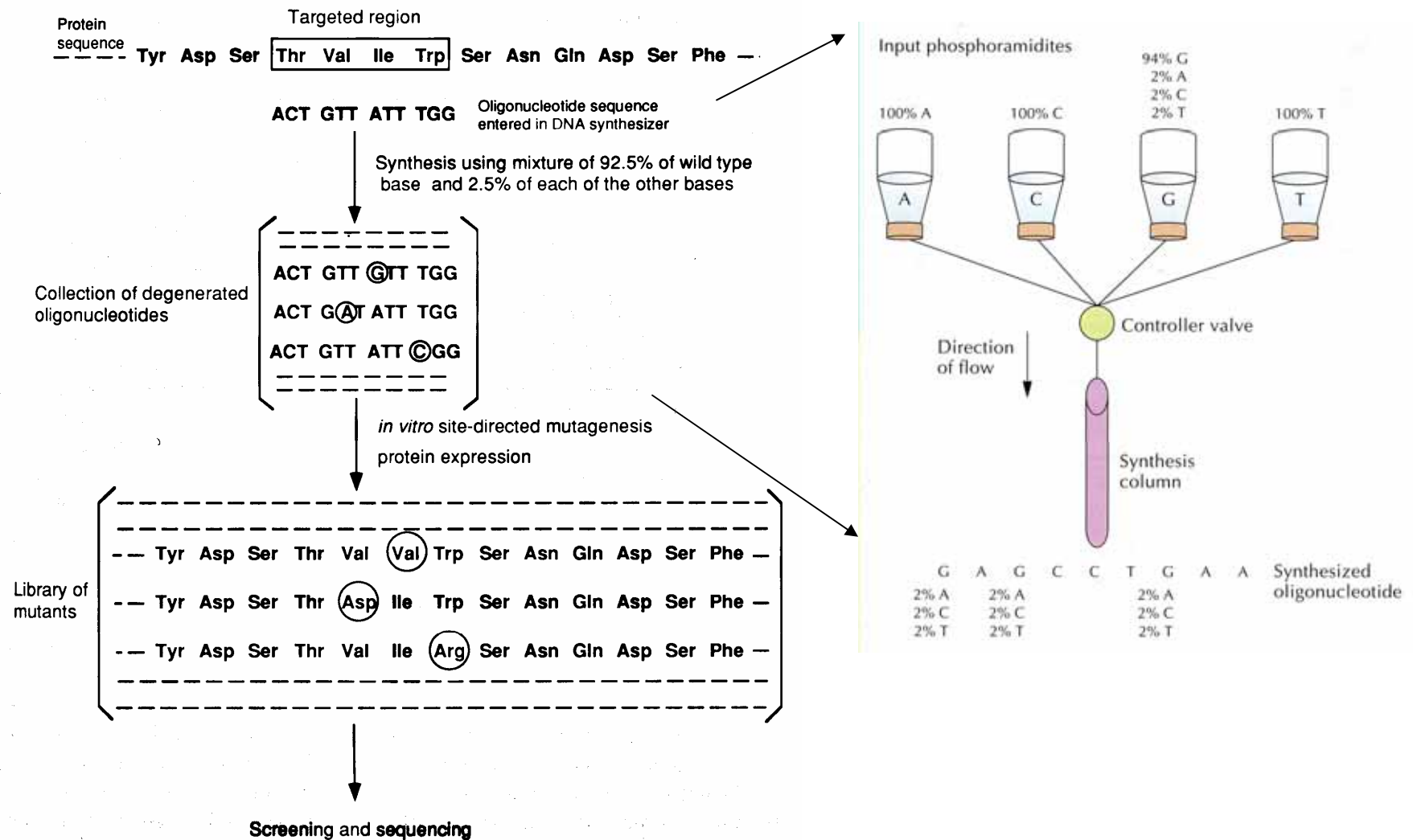
- ⇒ Beneficial mutations (good)
- ⇒ Neutral mutations
- ⇒ Deleterious mutations (bad)

⇒ Beneficial mutations are diluted with neutral and deleterious ones

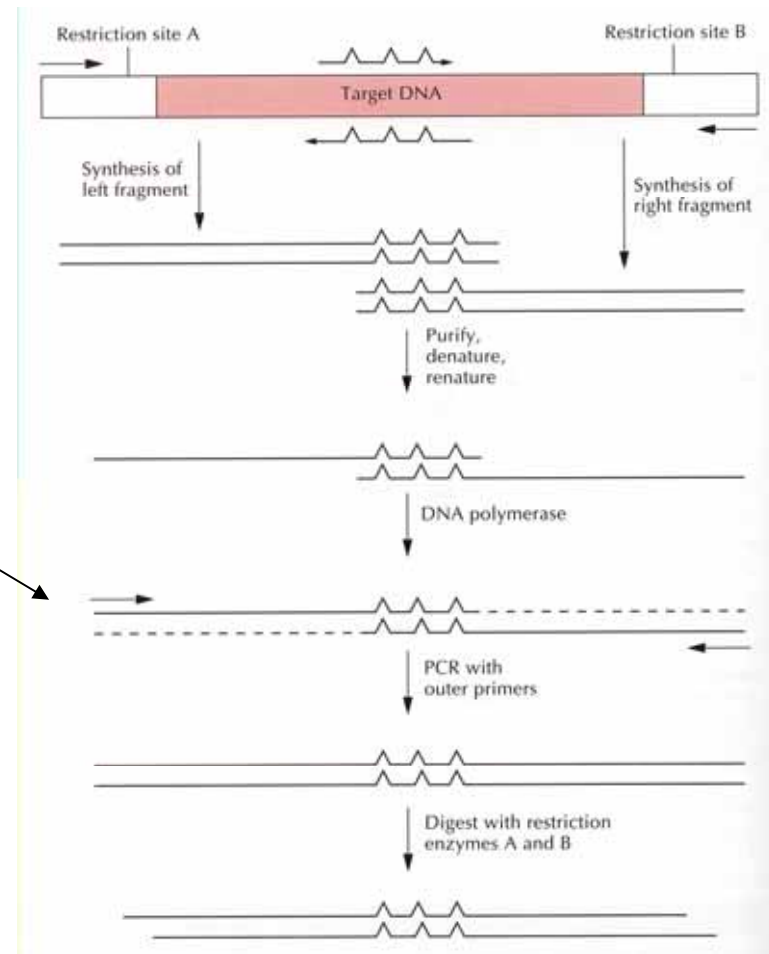
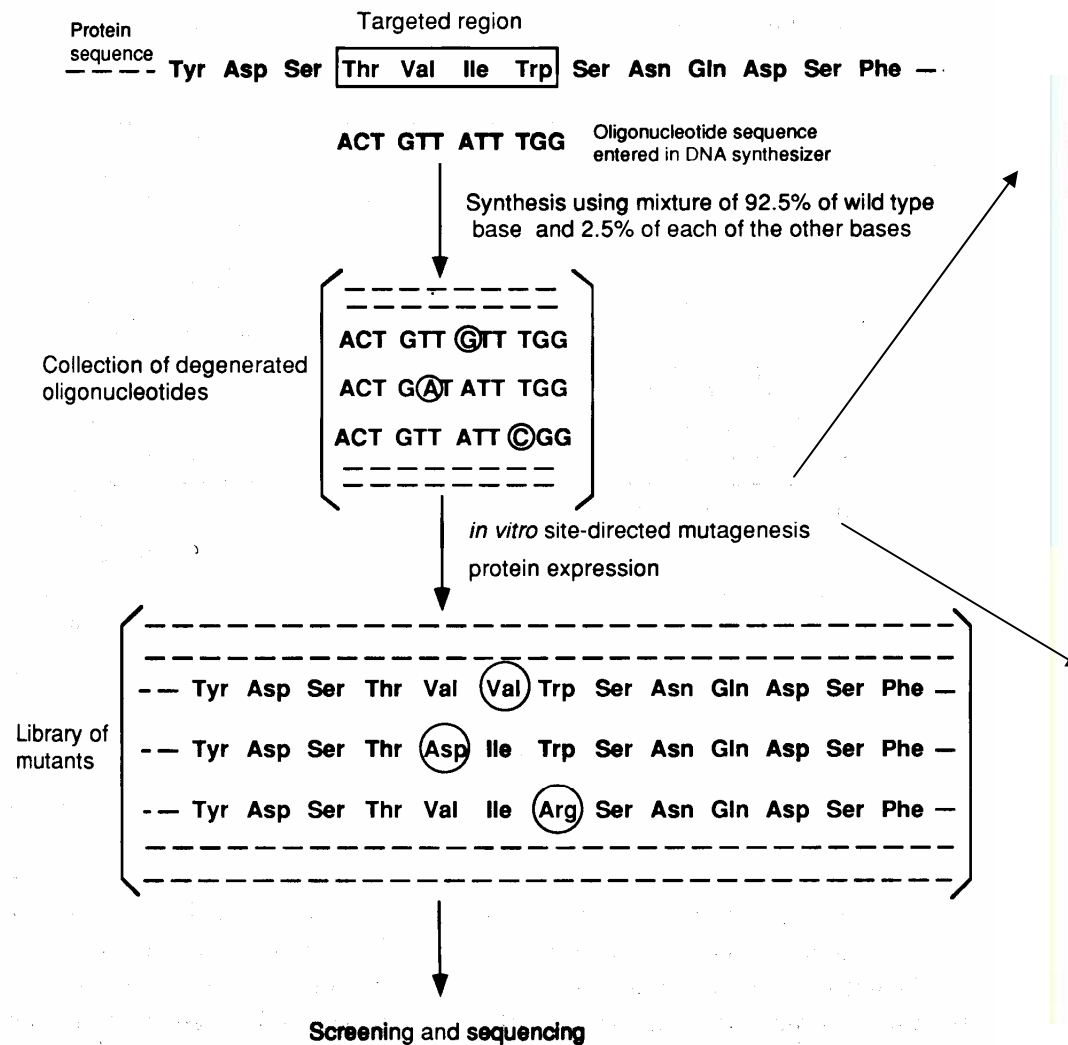
!!! Keep the number of mutations low per cycle

-> improve fitness of mutants!!!

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



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



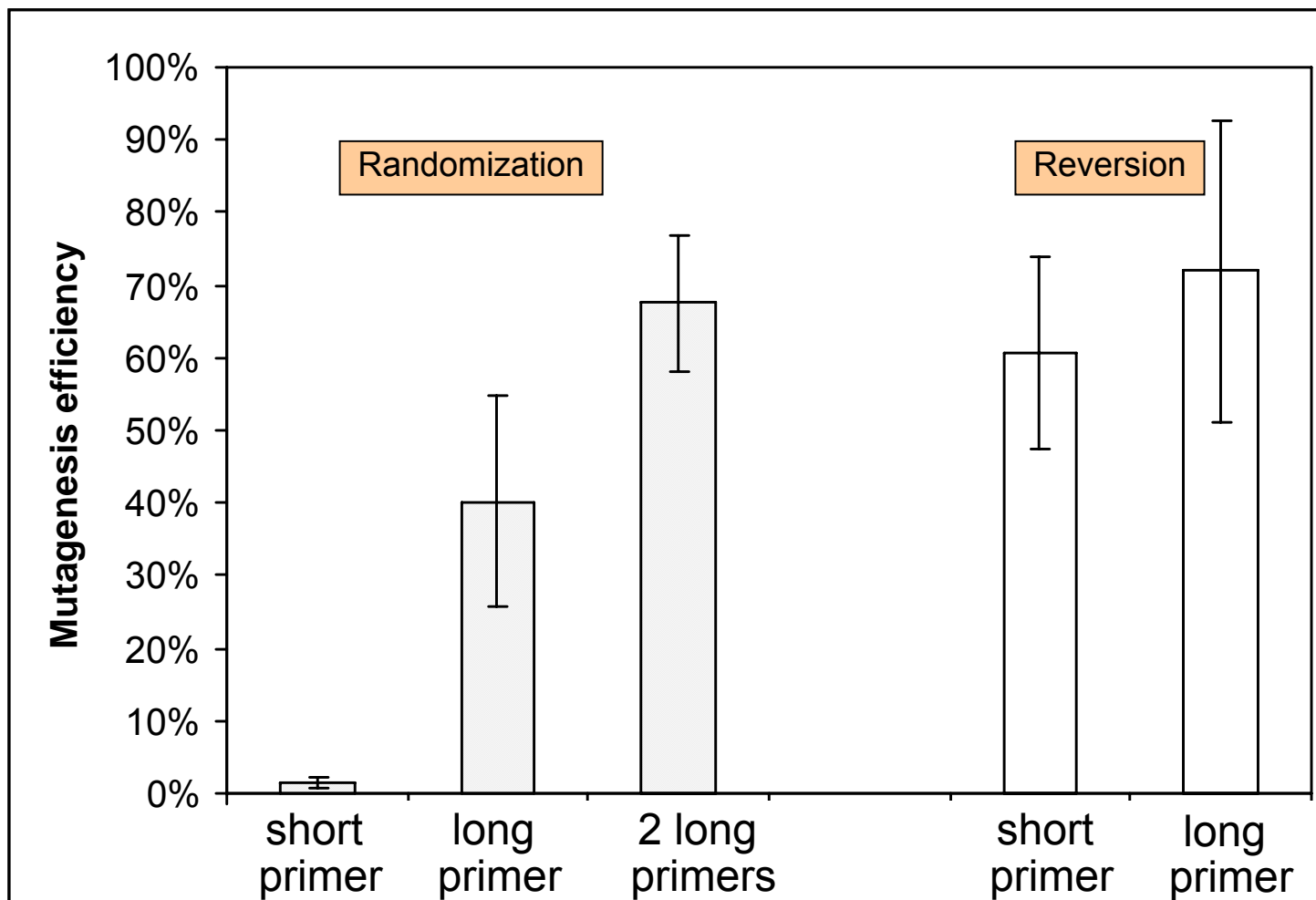
Site directed random mutagenesis of β -glucosidase Zm-p60.1 procedure evaluation

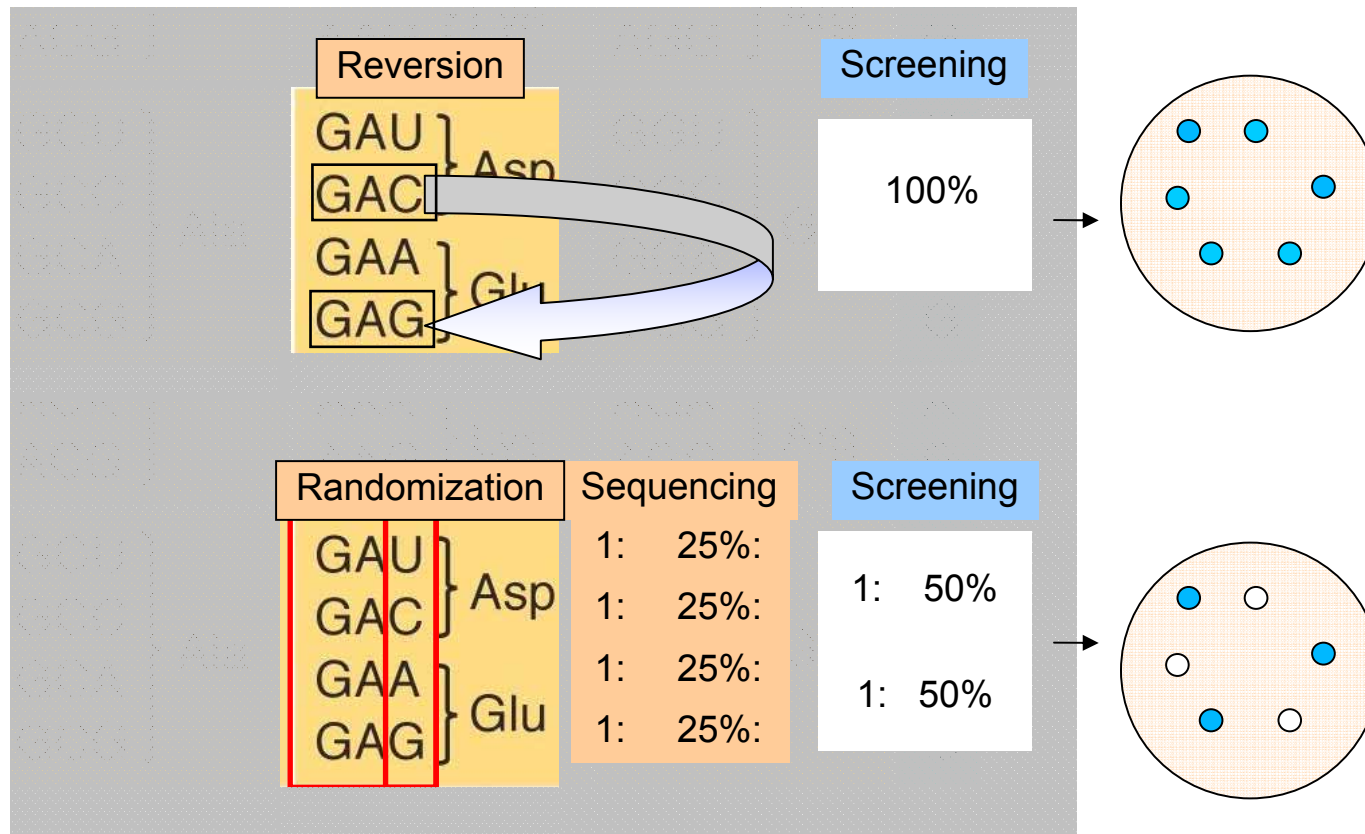
How much do theoretical genotype and phenotype frequencies
of mutants done by our random mutagenesis procedure differ from real frequencies?
(randomization accuracy or quality of mutant library)

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	<u>GAU } Asp</u> <u>GAC }</u> GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

Site directed random mutagenesis of β -glucosidase Zm-p60.1 procedure evaluation

		Second letter					
		U	C	A	G		
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	Third letter	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }		U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }		U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } GAC } Asp GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }		U C A G





Reversion

5'-gat tcc gtt Ctc cgt gat g -3' E401D → WT

5'-c ccc gat tcc gtt Ctc cgt gat gta gat agg tgg gtt tcc g -3' E401D → WT

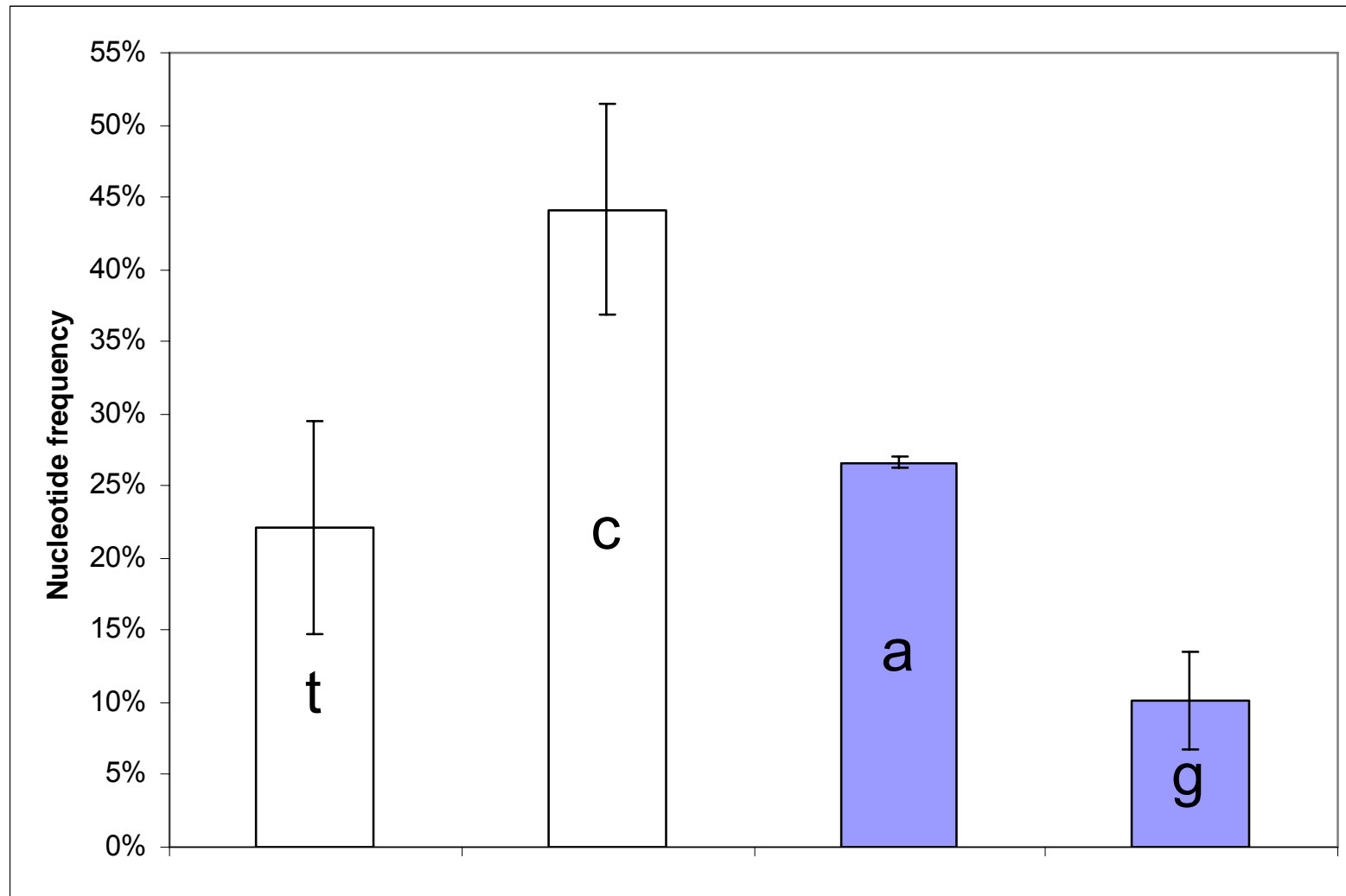
Randomization

5'-gat tcc gtt Ntc cgt gat g -3' E401D → VARIANTS

5'-c ccc gat tcc gtt Ntc cgt gat gta gat agg tgg gtt tcc g -3' E401D → VARIANTS

Additional primer complementary to the long E401D → VARIANTS.

5'-c gga aac cca cct atc tac atc acg gaN aac gga atc ggg g-3' E401D → VARIANTS



Observed frequencies of nucleotides at the mutated third position in codon 401.

Blue color depicts functional protein variants carrying these nucleotides while white color corresponds to catalytically inactive mutants.

- We expected a substantial bias of template sequences (cytosine) and the observed frequency was ~ 44%, which to a substantial extent explains the prevalence of white transformants.
- The frequencies of thymine and adenine were in the expected range (~ 25%).
- A surprising result was the frequency of guanine ~ 10%. This decrease could not be explained by the different properties of mutagenic primers. The calculated parameters like T_m , dimer and duplex forming energies do not show significant differences.

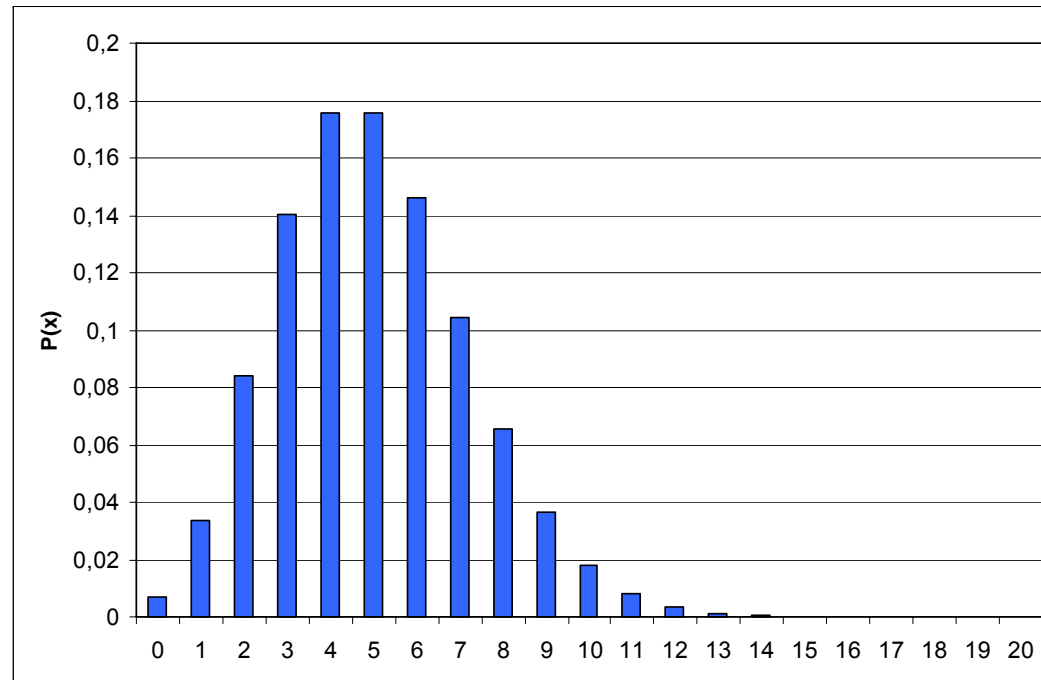
Theoretical implications for mutant library construction (size of the library)

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$\lambda = \frac{L}{V}$$

number of transformants (20)

number of variants (4)



Theoretical implications for mutant library construction (size of the library)

size of a library containing for example 95% of all possible variants

$$L = -V \ln 0.05 \approx 3V$$

$$L = -V \ln 0.05 K$$

size of a library with defined chance of being complete

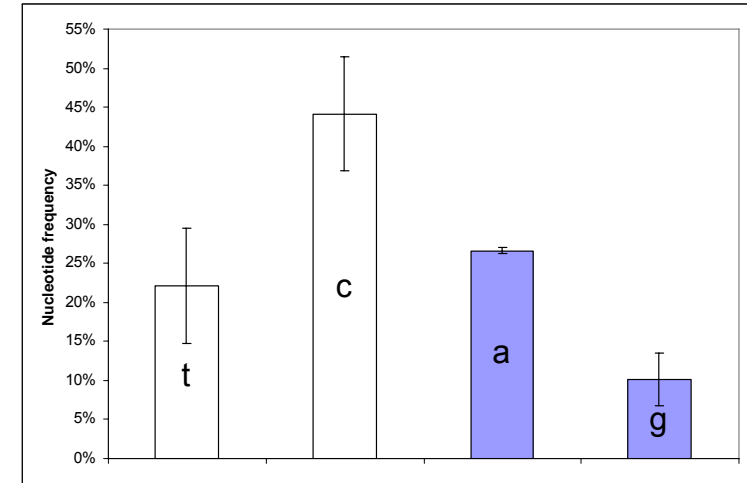
$$L = -V \ln \left[\frac{\ln P_c}{V} \right]$$

$$L = -V \ln \left[\frac{\ln P_c}{V} K \right]$$

P_c = probability that every variant is represented

L = number of transformants

V = number of variants

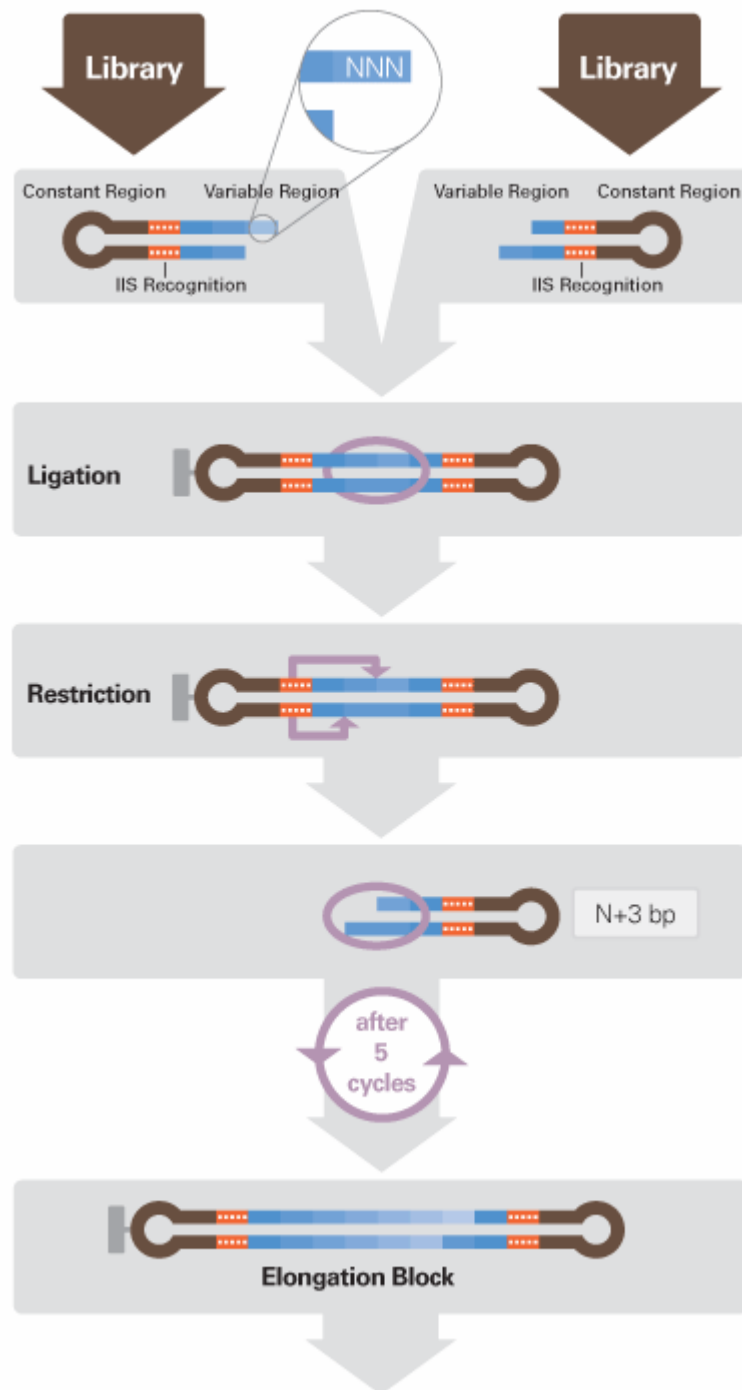


Theoretical implications for mutant library construction (size of the library)

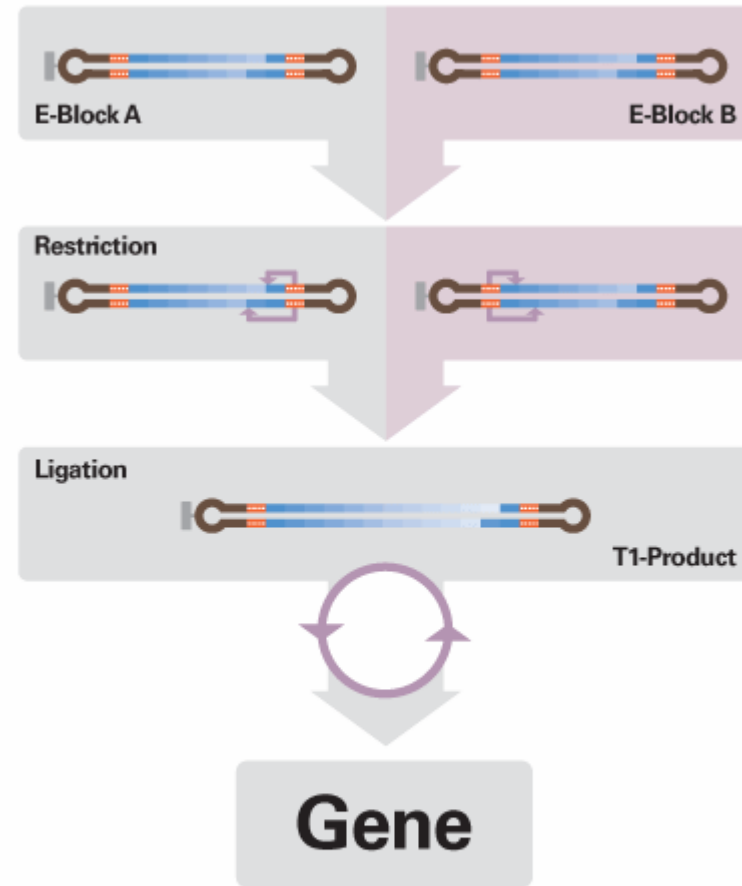
Mutant library parameters calculated for our sample via standard equations and modified by introducing empirical coefficient.

Number of transformants (L)=20 and
number of possible sequence variants (V)=4.

	Standard	Modified
Library size containing 95% of variants	12	53
Library size with 95% chance of being complete	17.4	77

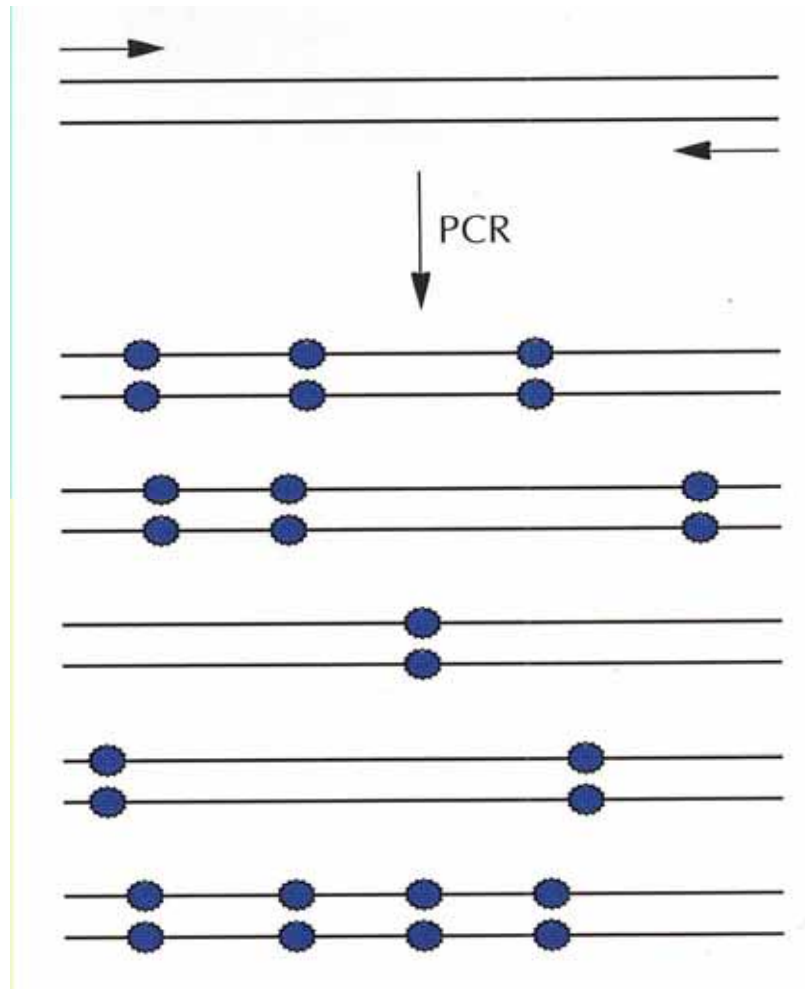


Slonmax



Random Mutagenesis (PCR based)

Error -prone PCR



-> PCR with low fidelity !!!

Achieved by:

- Increased Mg^{2+} concentration
- Addition of Mn^{2+}
- Not equal concentration of the four dNTPs
- Use of dI TP
- Increasing amount of Taq polymerase (Polymerase with NO proof reading function)

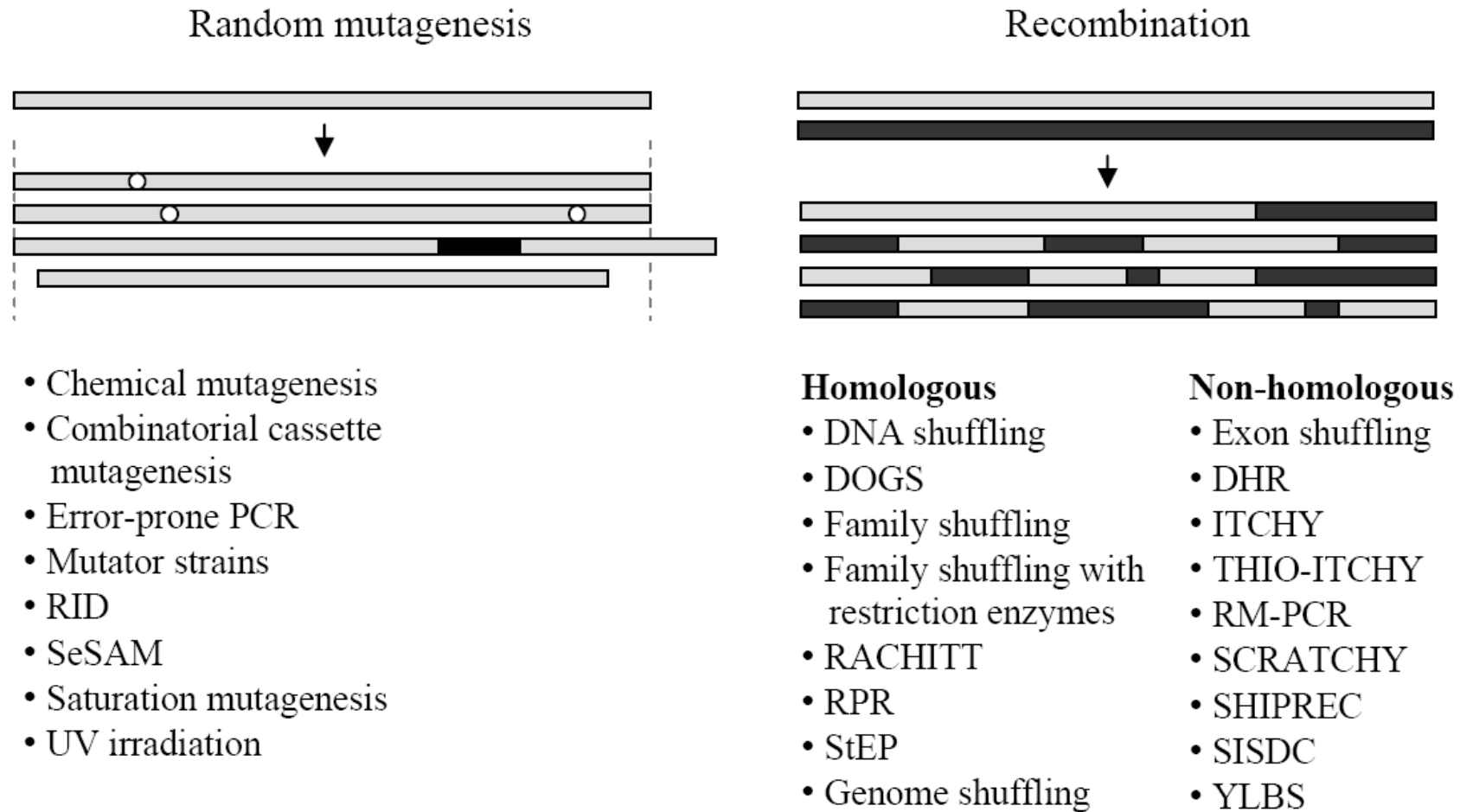
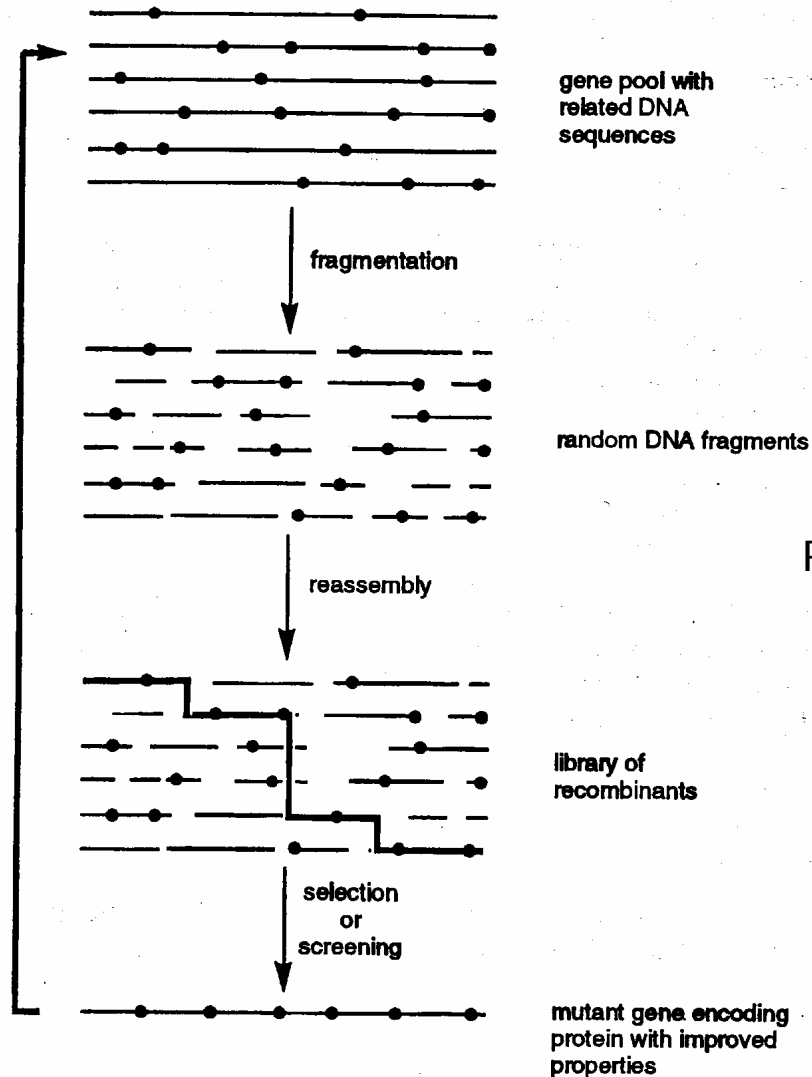


Fig. 1. Comparison of (a) random mutagenesis and (b) recombination strategies.

Random Mutagenesis (PCR based) DNA Shuffling

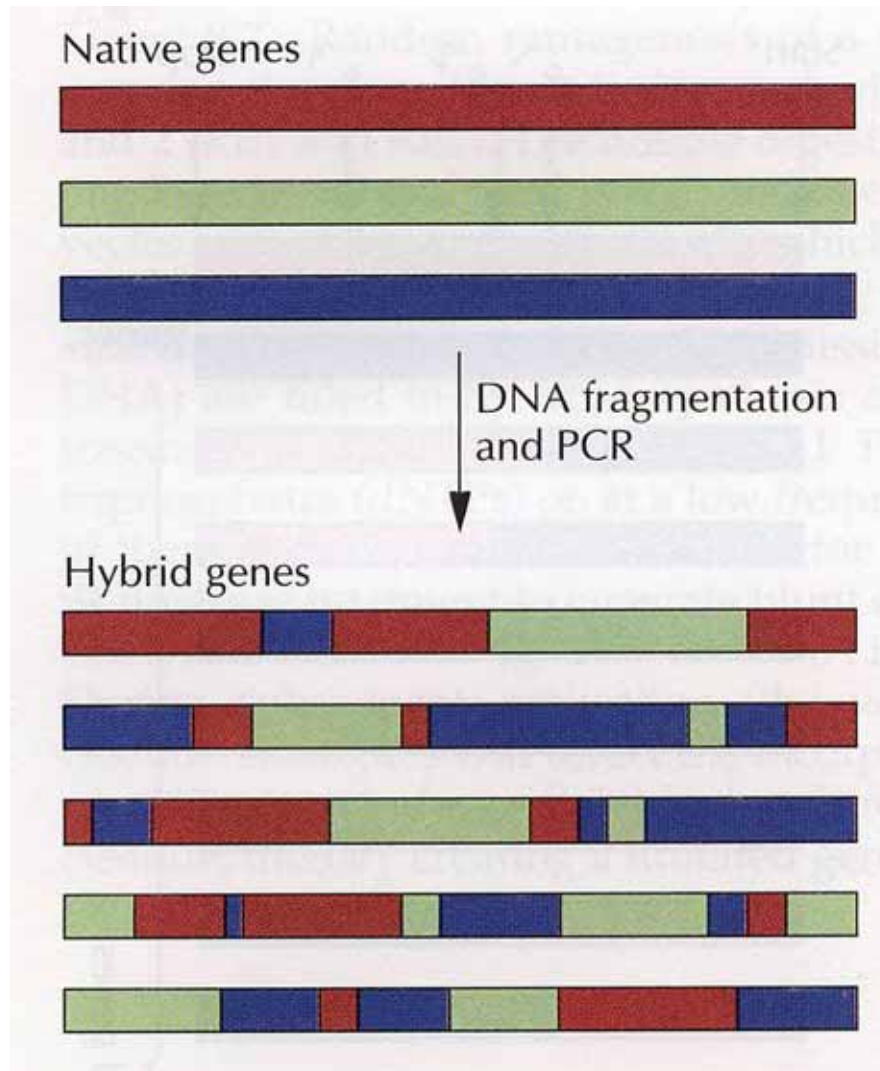


DNase I treatment (Fragmentation,
10-50 bp, Mn^{2+})

Reassembly (PCR without primers,
Extension and Recombination)

PCR amplification

Random Mutagenesis (PCR based) Family Shuffling



Genes coming from the same
gene family -> highly
homologous

-> Family shuffling

Random Mutagenesis (PCR based)

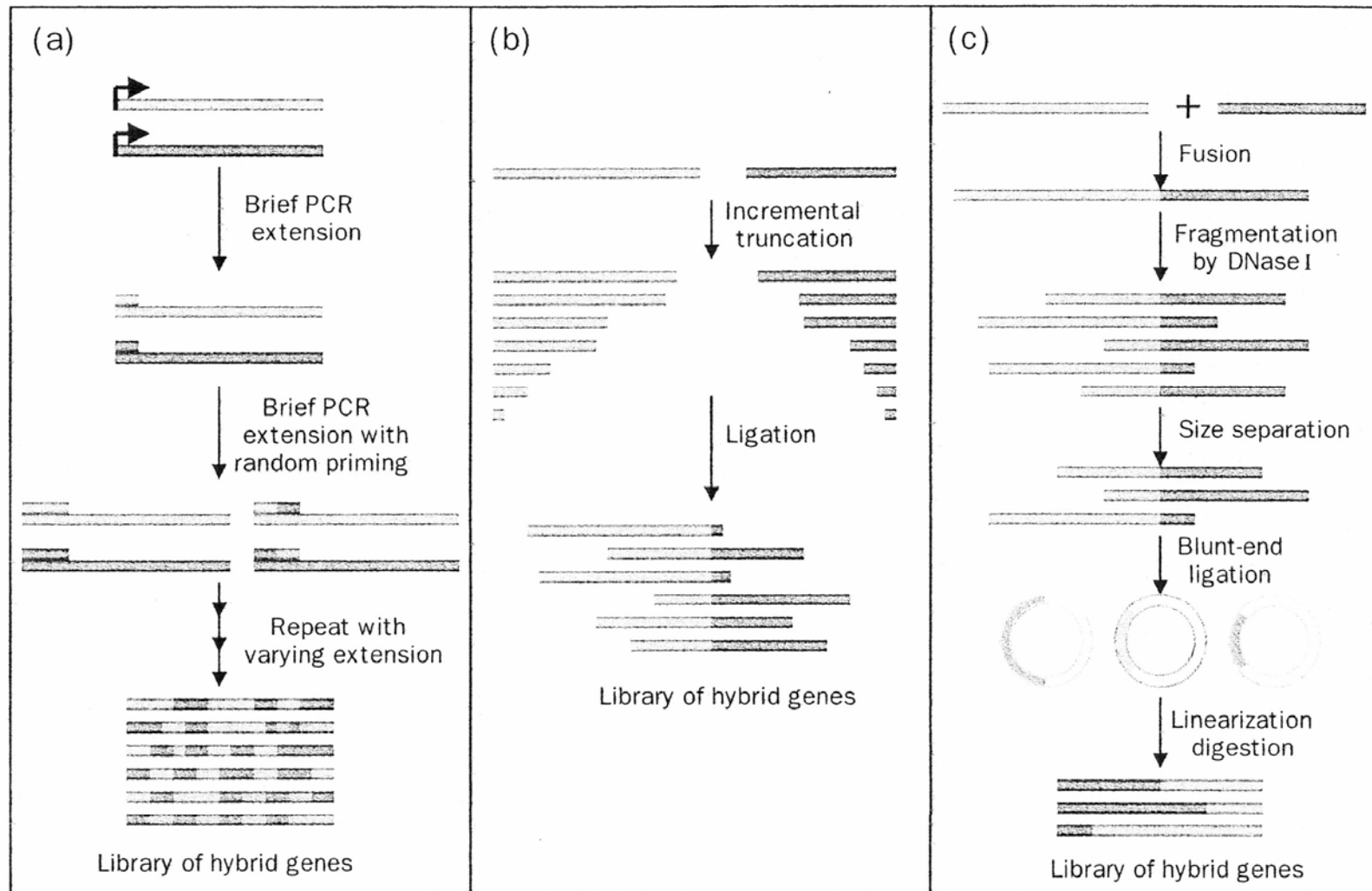


Figure 11 Procedures for the (a) StEP, (b) the ITCHY, and (c) the SHIPREC methods.

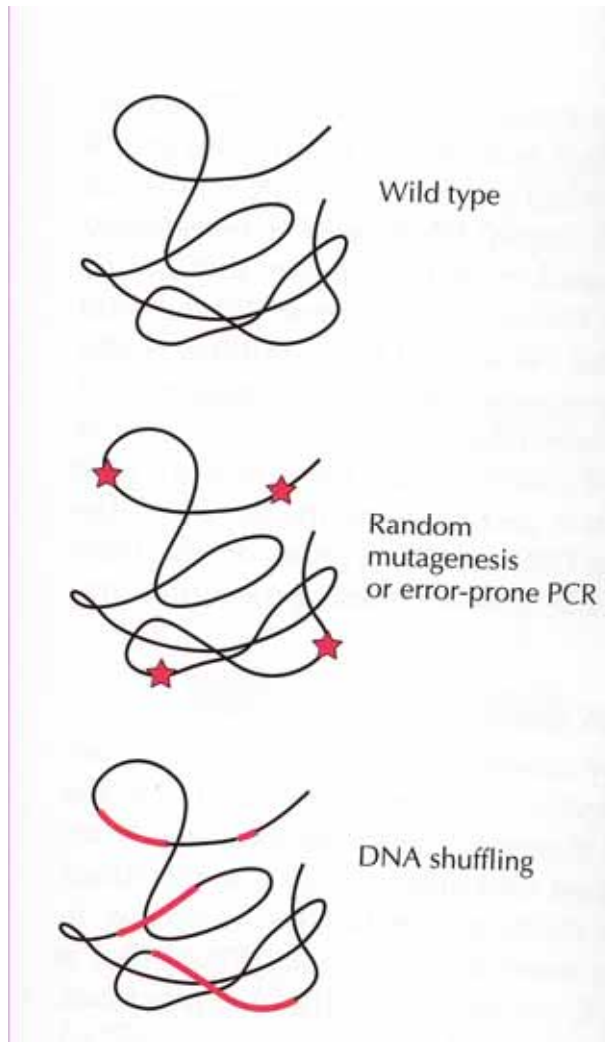
Staggered Extension Process (StEP)

Truncation for the Creation of Hybrid Enzymes (ITCHY)

Sequence Homology-Independent Protein Recombination (SHIPREC)

Directed Evolution

Difference between non-recombinative and recombinative methods



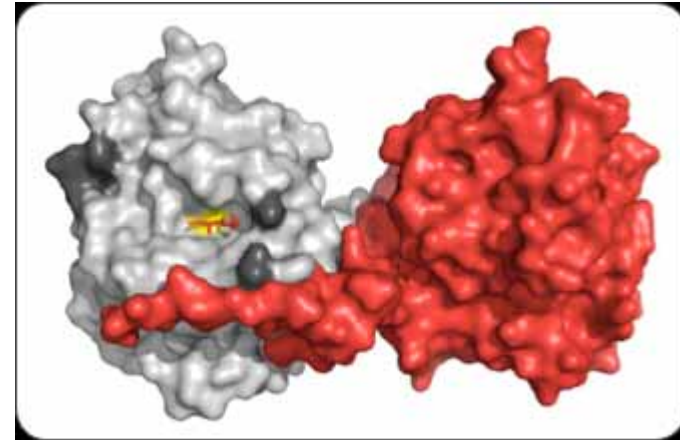
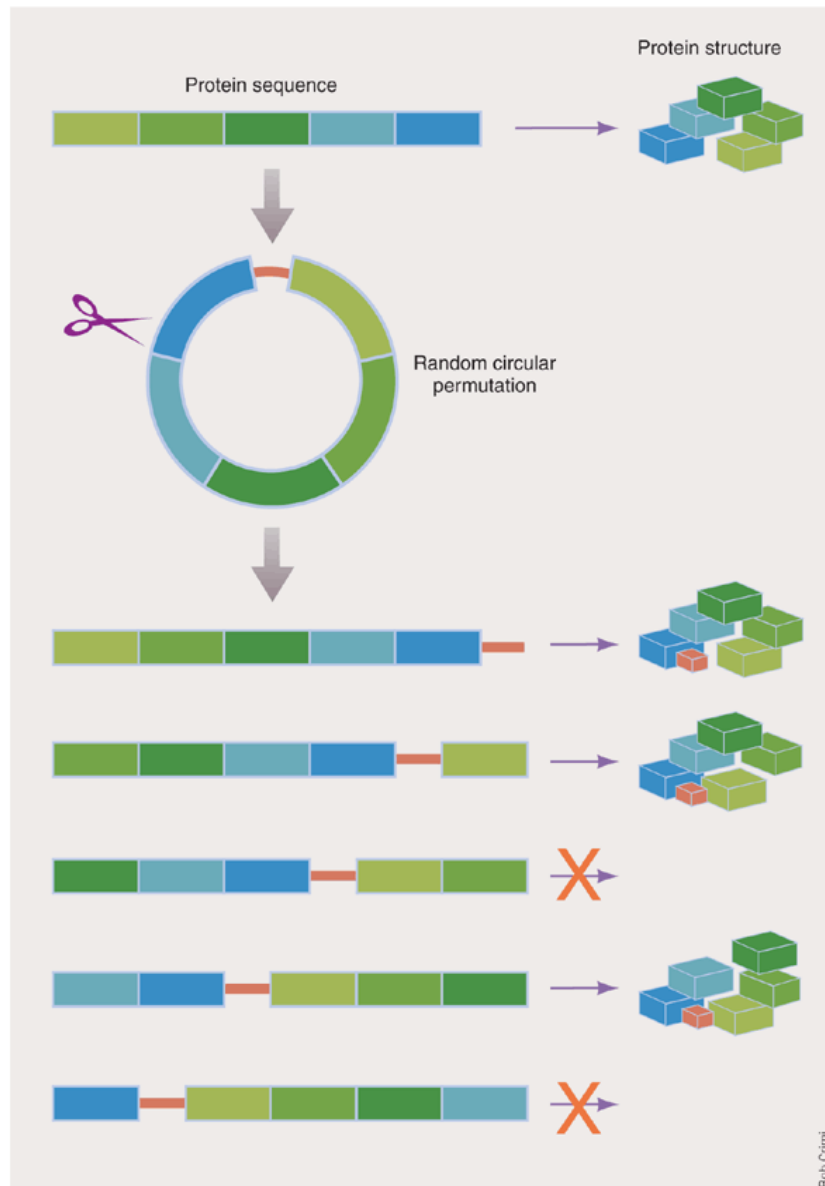
Non-recombinative methods

recombinative methods ->
hybrids (chimeric proteins)

...engineering proteins by circular permutation

Protein engineering must not necessarily involve the substitution of amino acids. The reorganization of a proteins' primary sequence can also change the catalytic properties. We are using a technique called circular permutation to explore the effects of termini relocation on catalysis, as well as protein stability & dynamics.

...engineering proteins by circular permutation



lipase B from *Candida antarctica*, an important biocatalyst in asymmetric synthesis. Upon relocating the protein termini by circular permutation, we observe up to 175-fold enhanced catalytic performance while preserving the enzyme's enantio-selectivity.

Stefan Lutz group

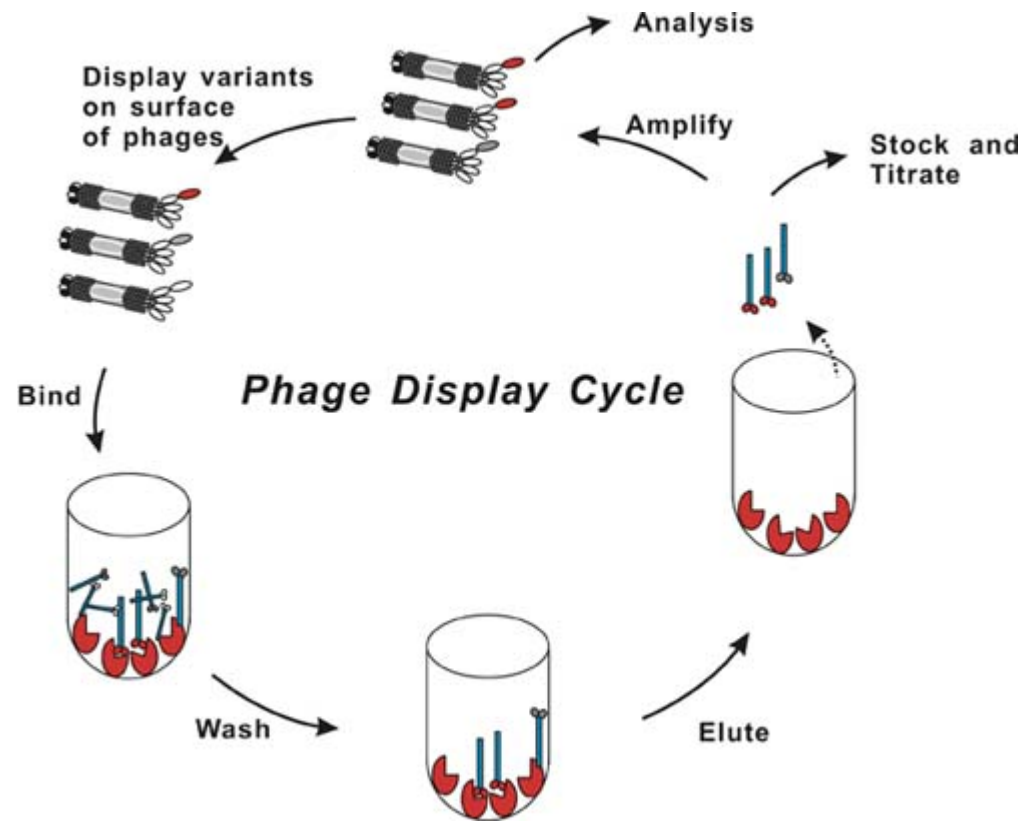
HTS (High throughput screening) System

Screening methods

- **Genetic selection**
 - **Growth**
 - **Survival**
- **Display technology**
 - **In-vitro display (cell-free translation)**
 - **Phage display**
 - **Cell surface display (Bacteria & Yeast)**
- **Solid or liquid-phase assay**

Screening systems

- **FACS (Fluorescence-activated cell sorter)**
- **Digital image spectroscopy**
- **Fluorescence detection technique**



HTS (High throughput screening) System

Screening methods

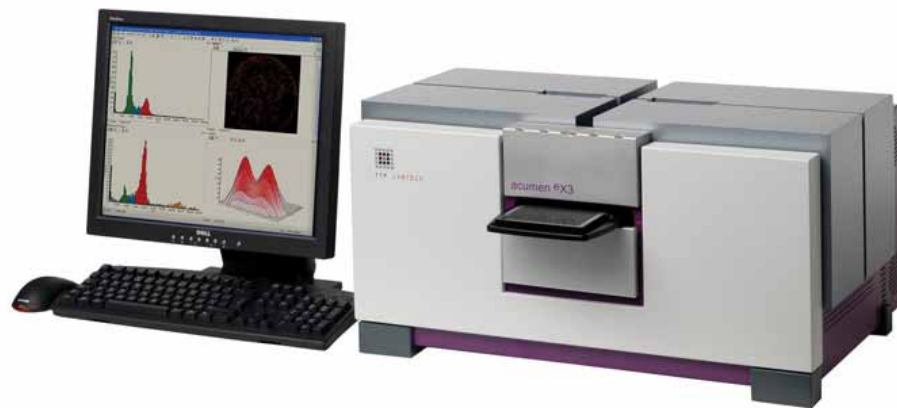
- **Genetic selection**
 - **Growth**
 - **Survival**
- **Display technology**
 - **In-vitro display (cell-free translation)**
 - **Phage display**
 - **Cell surface display (Bacteria & Yeast)**
- **Solid or liquid-phase assay**

Screening systems

- **FACS (Fluorescence-activated cell sorter)**
- **Digital image spectroscopy**
- **Fluorescence detection technique**

High Content Analysis and High Content Screening

High Content Analysis (HCA) and High Content Screening (HCS) are imaging based multi-parametric approaches of cell analysis at the single-cell level. Originally developed as a complementary technology to traditional biochemical high-throughput screening (HTS) in drug discovery, today High Content Screening is established in a far broader area of the life science space as an unbiased method of imaging multiple cellular samples.

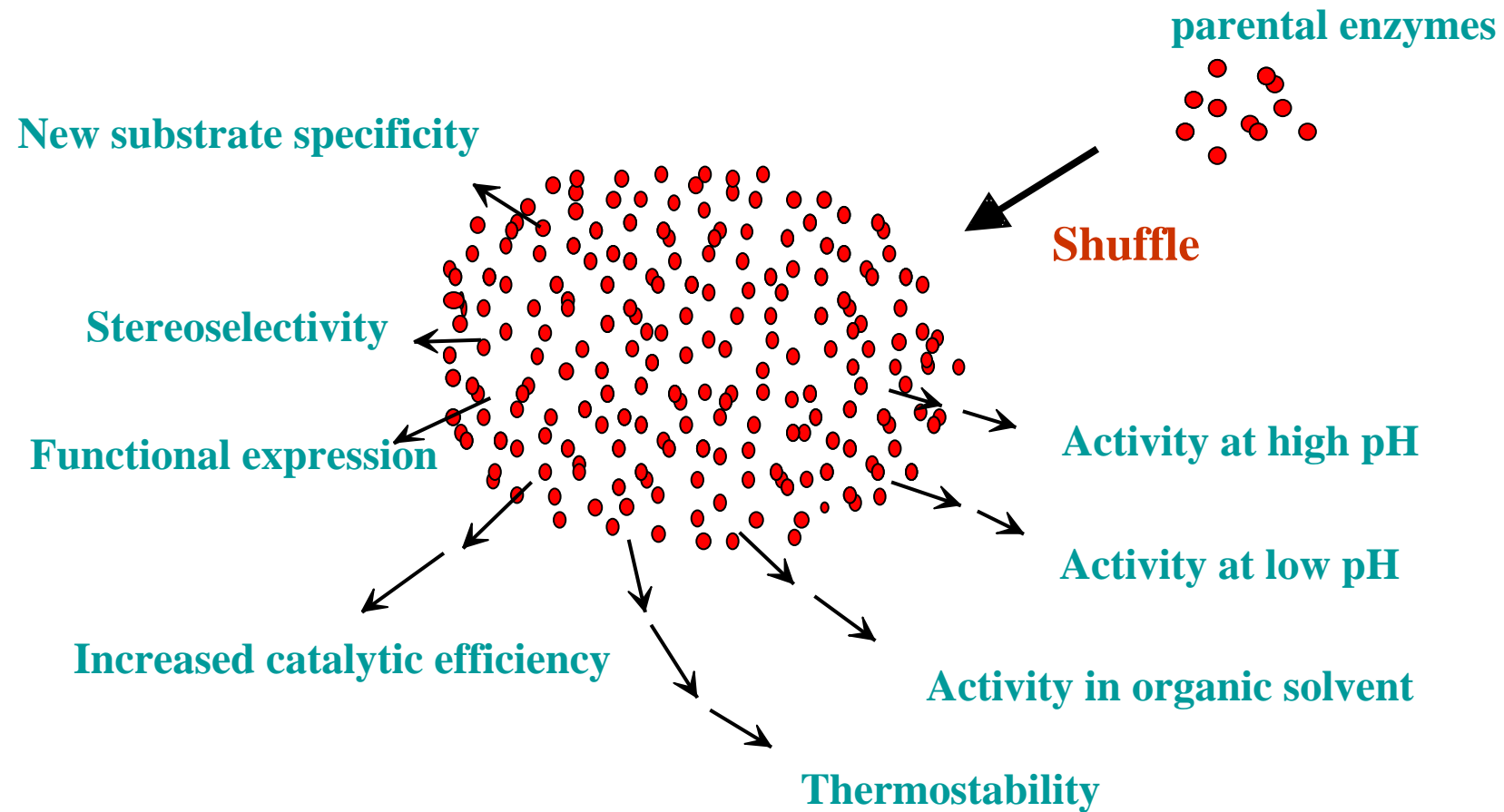


Miami Project High Content Screening Core

The core technology in the HCS uses the Cellomics VTI ArrayScan with a Thermo Fisher robot to feed the scanner.

This system can scan about thirty 96 well plates over night, acquiring 9 fields per well, with 3-4 colors. This is a lot of data.

Functional Diversity



Protein Engineering - Applications

Engineering Stability of Enzymes – T4 lysozyme

-> S-S bonds introduction

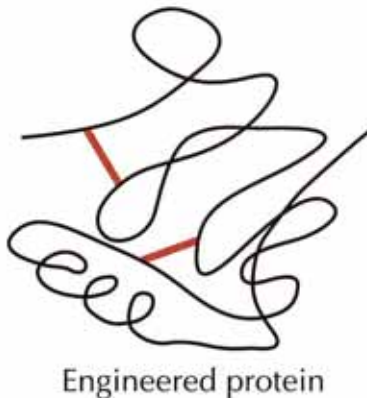


Table 8.2 Properties of T4 lysozyme and six engineered variants

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).

Protein Engineering - Applications

Engineering Stability of Enzymes – triosephosphate isomerase from yeast

-> replace Asn (deaminated at high temperature)

Table 8.3 Stability at 100°C of the yeast enzyme triosephosphate isomerase and its engineered derivatives

Enzyme	Amino acid at position:		Half-life (min)
	14	78	
Wild type	Asn	Asn	13
Variant A	Asn	Thr	17
Variant B	Asn	Ile	16
Variant C	Thr	Ile	25
Variant D	Asp	Asn	11

Adapted from Ahern et al., *Proc. Natl. Acad. Sci. USA* **84**:675–679, 1987.

Enzyme stability is expressed as the half-life, or rate of enzyme inactivation, at 100°C. A longer half-life indicates a more stable enzyme.

Protein Engineering - Applications

Engineering Activity of Enzymes – tyrosyl-tRNA synthetase from *B. stearothermophilus*

-> replace Thr 51 (improve affinity for ATP) -> Design

Table 8.4 Aminoacylation activity of native (Thr-51) and modified (Ala-51 and Pro-51) tyrosyl-tRNA synthetases

Enzyme	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($\text{s}^{-1} \text{M}^{-1}$)
Thr-51	4.7	2.5	1,860
Ala-51	4.0	1.2	3,200
Pro-51	1.8	0.019	95,800

Adapted from Wilkinson et al., *Nature* **307**:187–188, 1984.

The units for K_m , the binding constant of the enzyme for ATP, are millimolar units (mM); the units for k_{cat} , the catalytic rate constant, are reciprocal seconds (s^{-1}); and the units for k_{cat}/K_m , the catalytic efficiency, are $\text{s}^{-1} \text{M}^{-1}$.

Protein Engineering - Applications

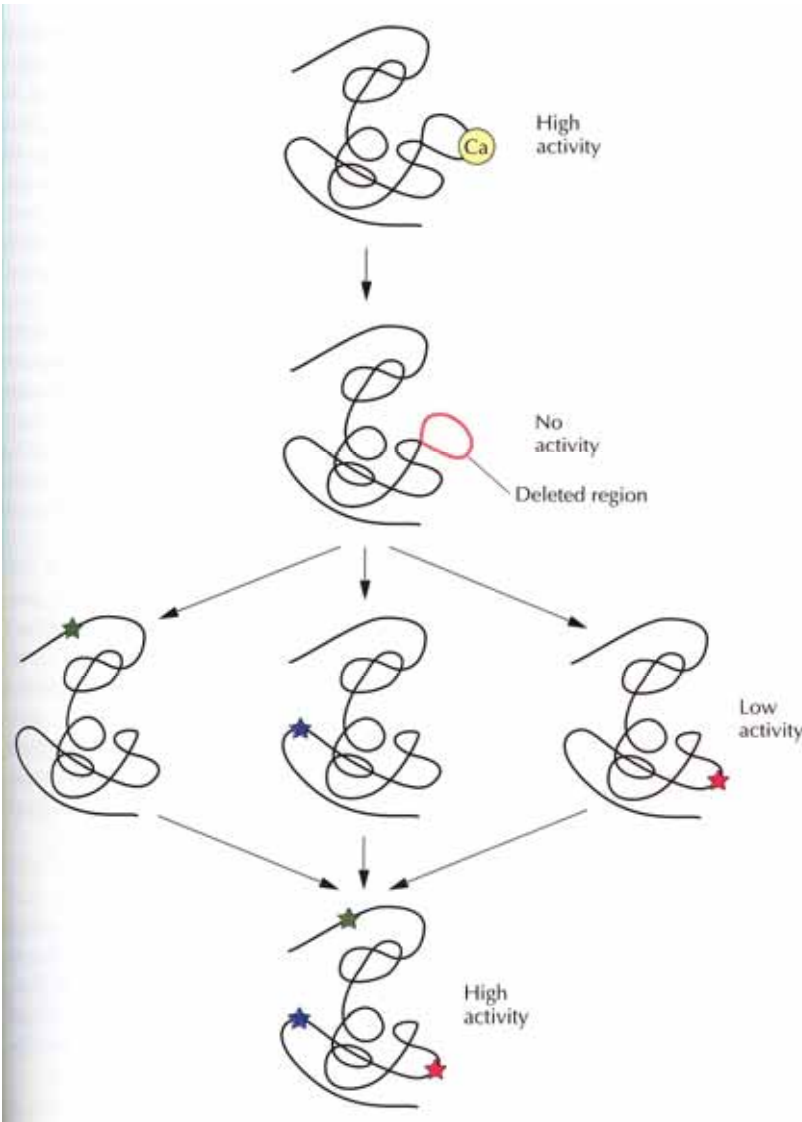
Engineering Ca-independency of subtilisin

Saturation mutagenesis -> 7 out of 10 regions were found to give increase of stability

Mutant:

10x more stable than native enzyme in absence of Ca

50% more stable than native in presence of Ca



Protein Engineering - Applications

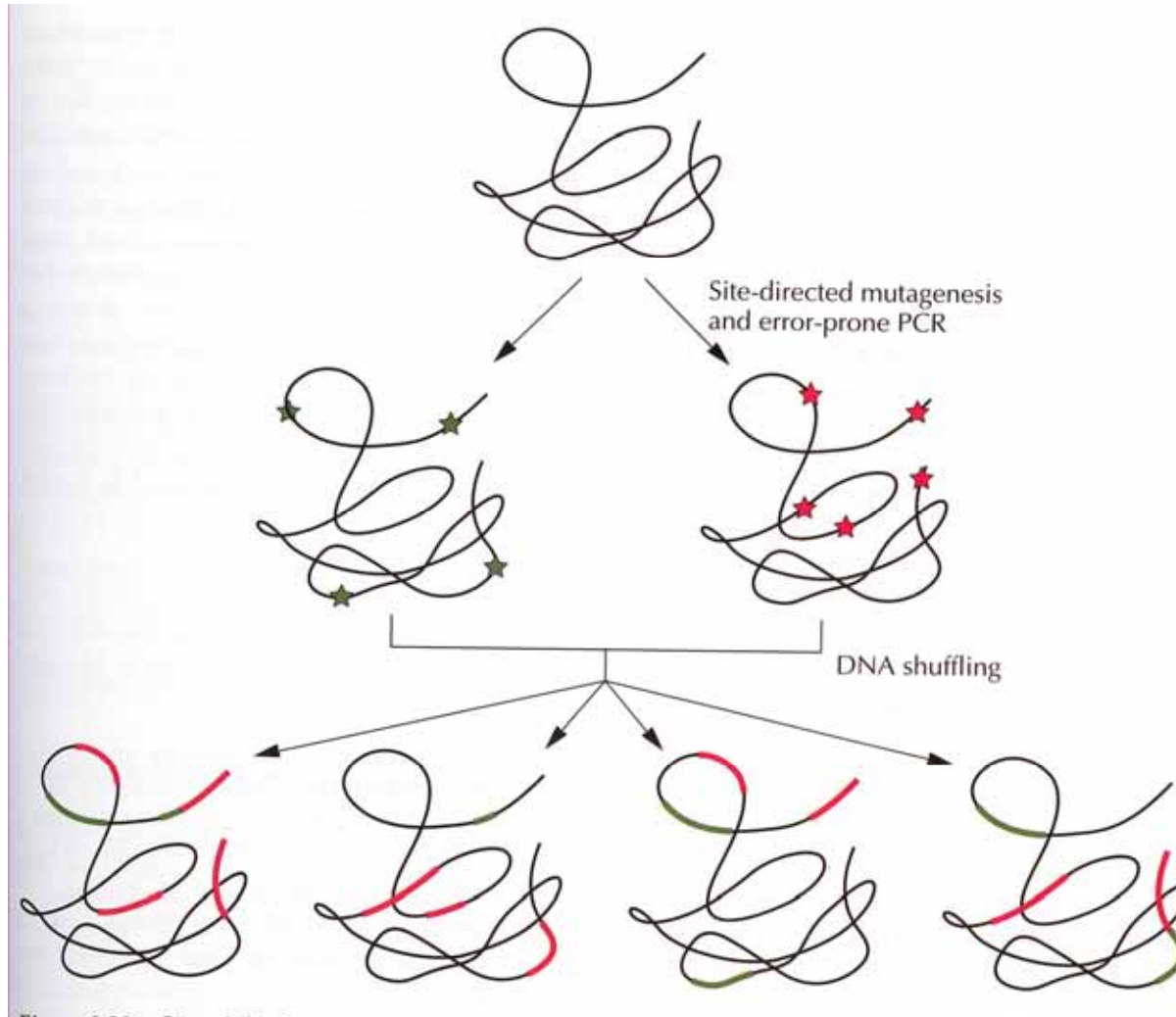
Site-directed mutagenesis -> used to alter a single property

Problem : changing one property -> disrupts another characteristics

Directed Evolution (Molecular breeding) -> alteration of multiple properties

Protein Engineering – Applications

Directed Evolution



Protein Engineering – Applications

Directed Evolution

Table 1. Examples of the diversity of properties that have been improved by evolutionary engineering

Protein	Altered function	Reference
Barley α -amylase	Thermostability: ten-fold increase of half-life at 90°C	JOYET et al. 1992
Subtilisin	Alkaline stability: doubling the autolytic half-time at pH 12	CUNNINGHAM and WELLS (1987)
Subtilisin	Tolerates loss of stabilizing divalent cations	STRAUSBERG et al. (1995)
Subtilisin E	Active in 60% DMF	YOU and ARNOLD (1996)
<i>Streptomyces griseus</i> protease B	Broadened substrate specificity	SIDHU and BORGFORD (1996)
Green fluorescent protein	40-fold brighter fluorescing bacterial colonies	CRAMERI et al. (1996)
Immunoglobulin constant domain	Preferential formation of heterodimers	ATWELL et al. (1997)
Immunoglobulin variable domain	Tolerates loss of structural disulfide bridge	MARTINEAU et al. (1998)

Protein Engineering – Applications

Directed Evolution

Table 2. Representative experiments using successive cycles of variation and selection

Protein	Property	Number of cycles for success	Number of nucleotide changes required	Number of amino acid changes required	Reference
β -Lactamase	Increased activity	3	4	4	STEMMER (1994b)
GFP	Improved folding and expression	3	3	3	CRAMERI et al. (1996)
Subtilisin E	Stability in aqueous DMF	2	3	3	YOU and ARNOLD (1996)
Arsenite membrane pump	Increased activity	3	3	3	CRAMERI et al. (1997)
FLP-recombinase	Thermostability	8	3–4	3–4	BUCHHOLZ et al. (1998)

In almost all cases a single nucleotide change leading to a single amino acid change was sufficient per cycle, the number of silent mutations was approximately the same. No amino acid change was reported that would have required more than one nucleotide change. Thus current protocols appear to sample sequence space in a biased fashion, in single mutation steps.

GFP

