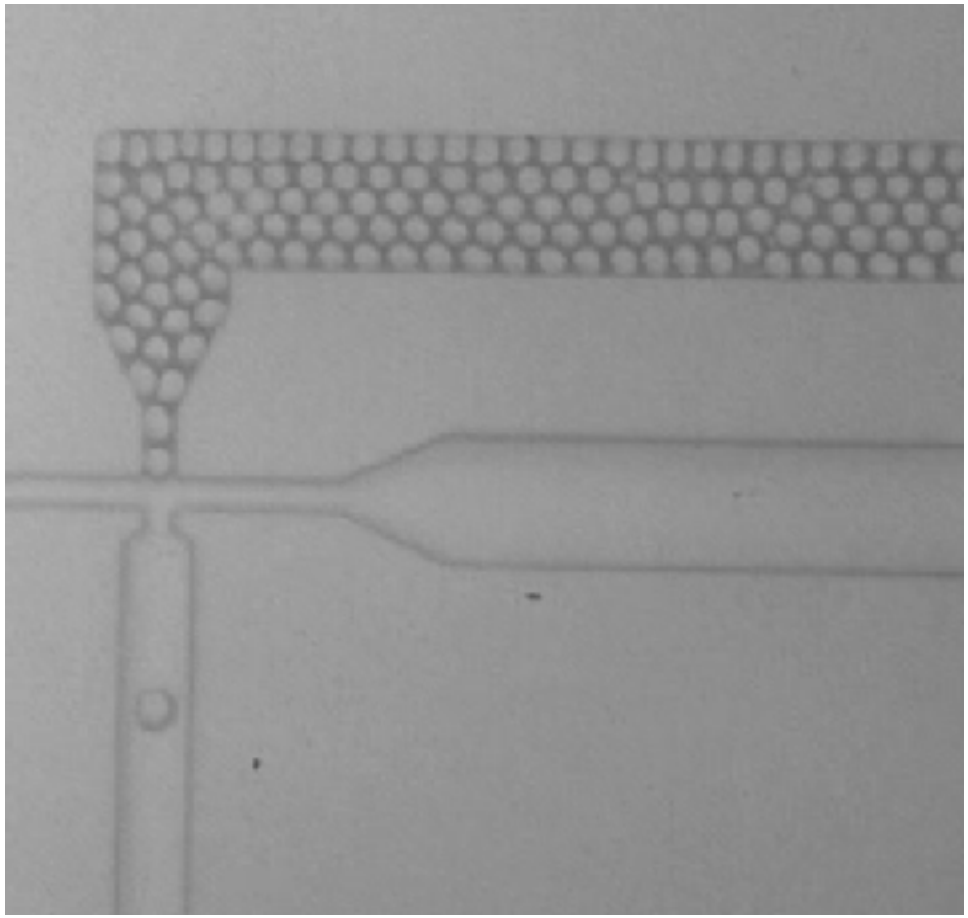
Biotechnologie Moléculaire

3ème année

Andrew Griffiths

Laboratoire de Biochimie

Room B 231
Tel: 01 40 79 45 39
Email: andrew.griffiths@espci.fr



Biotechnologie Moléculaire

3ème année

The lectures will be available on the intranet:

https://cours.espci.fr/site.php?id=258

Other information relative to the course will also be posted here:

Engineering Proteins by Design

Modelling protein structures and interactions

Engineering Proteins by Design

Modelling of protein structures and interactions

Some important milestones in protein design have been achieved recently:

1. In design of *de novo* folded proteins

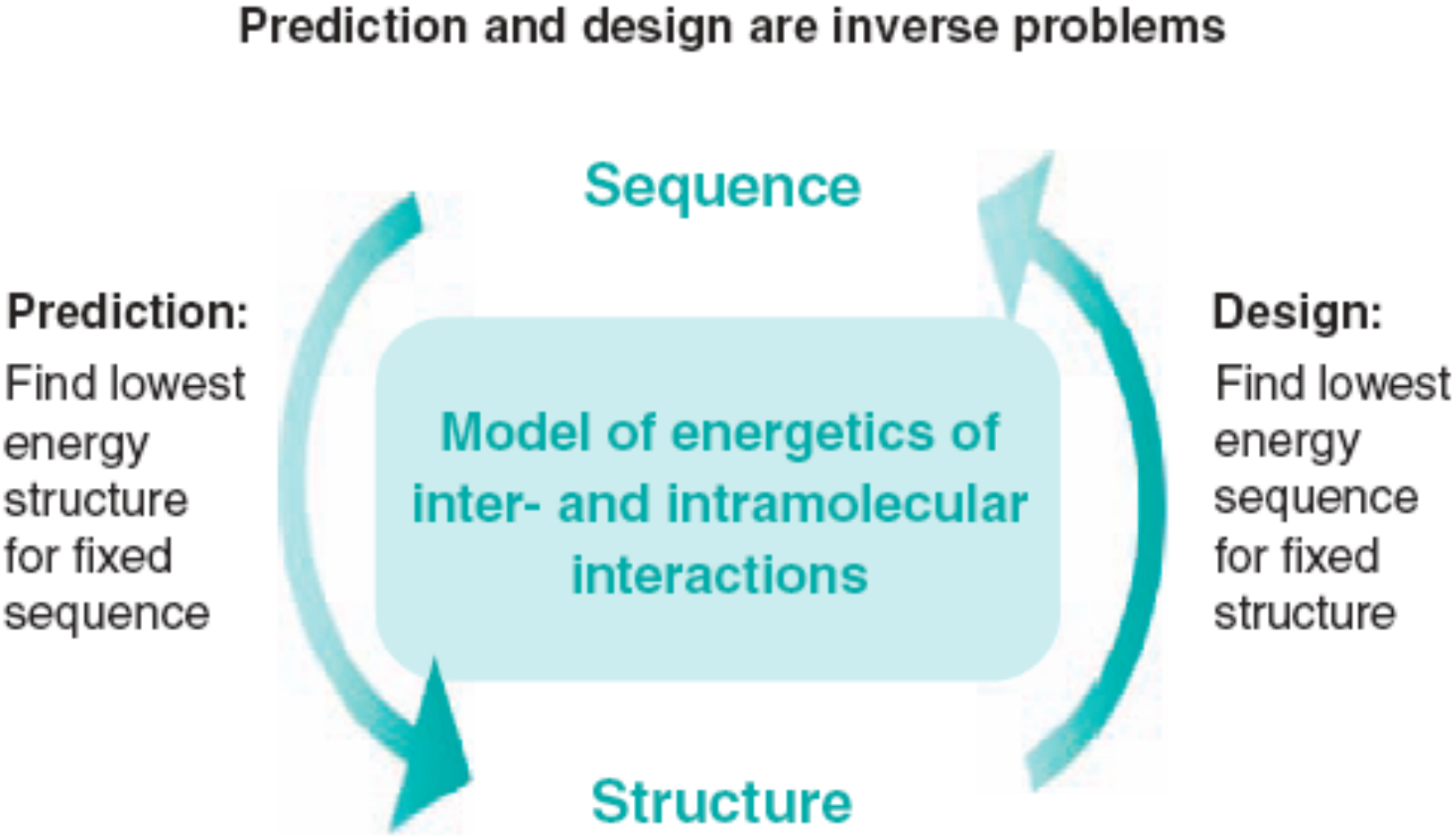
2. In design of existing proteins with new functions

There is a huge future potential.

However, to date design has had only a small impact on engineering of therapeutic proteins

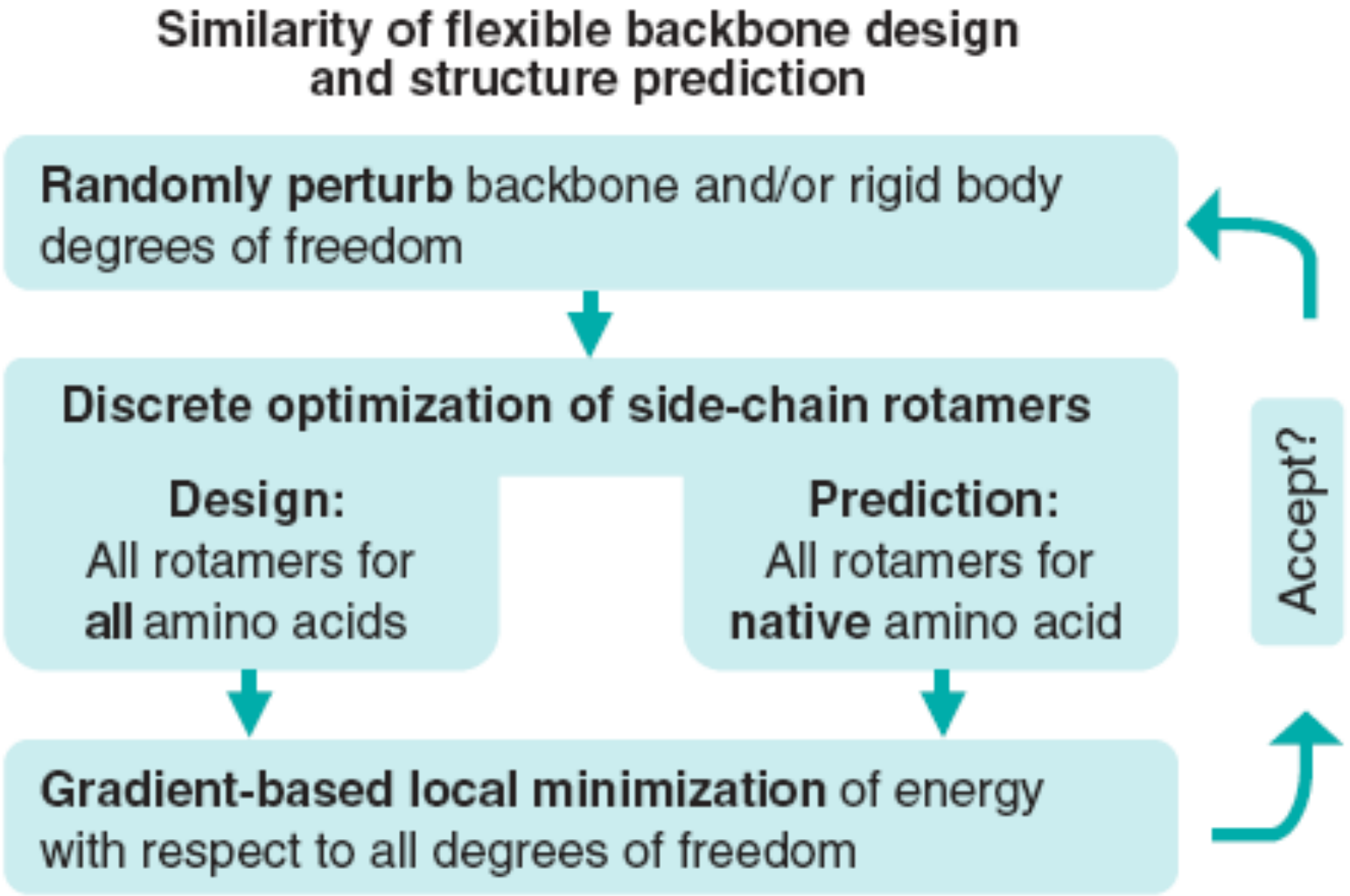
Engineering Proteins by Design

Prediction versus design



Engineering Proteins by Design

Monte Carlo minimization (MCM) high-resolution refinement protocol



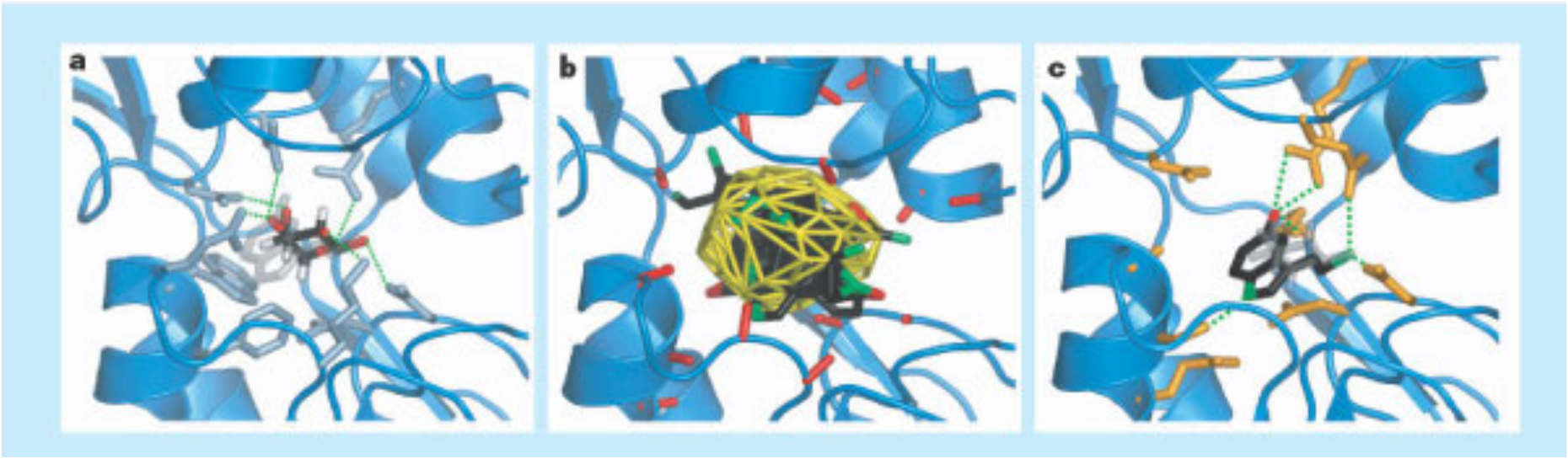
The same basic MCM algorithm can be used for structure prediction, protein-protein docking, and flexible backbone design

Engineering Proteins by Design

Design of a receptor with new ligand-binding activities

Design of a receptor which causes bacteria to turn green when exposed to trinitrotoluene (TNT)

Periplasmic binding proteins (PBPs) normally signal to the cell on binding arabinose
A PBP was re-designed to bind TNT, serotonin or L-lactate



wt PBP binding arabinose. Complementary surface of receptor defined (pale blue)

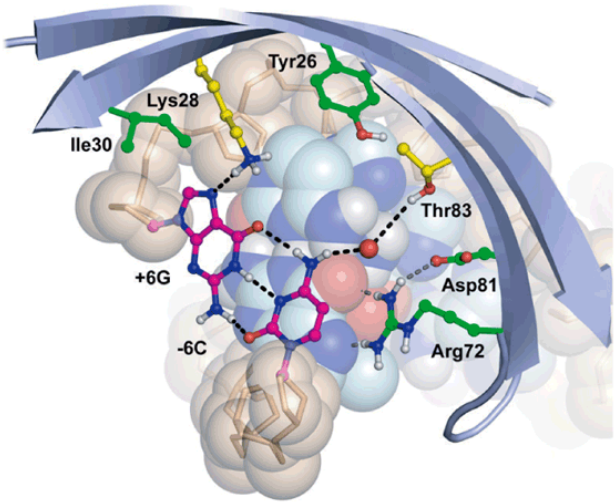
Amino acids in complementary surface truncated to alanine. Serotonin molecules docked in confined convex polygonal space (yellow) in cavity

New side chains tested and serotonin docked defining new complementary surface (gold)

Looger, L.L., Dwyer, M.A., Smith, J.J. and Hellinga, H.. *Nature* 423, 185-190 (2003).

Engineering Proteins by Design

Redesign of endonuclease DNA binding and cleavage specificity



X-ray structure of wild-type
I-*Mso*I homodimer complex

Consider (in turn) all symmetric base pair changes

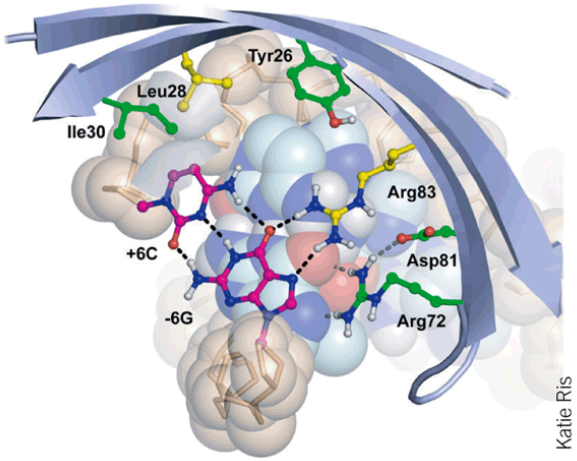
Try new side chains near these base pairs

List predicted discrimination energy

Predict best altered discrimination with symmetric G-C changes at +6 and –6 and with K28L & T83R changes in each monomer

Synthesize and test this variant: excellent cleavage rate and discrimination at new site

X-ray structure of variant confirms the predicted contacts

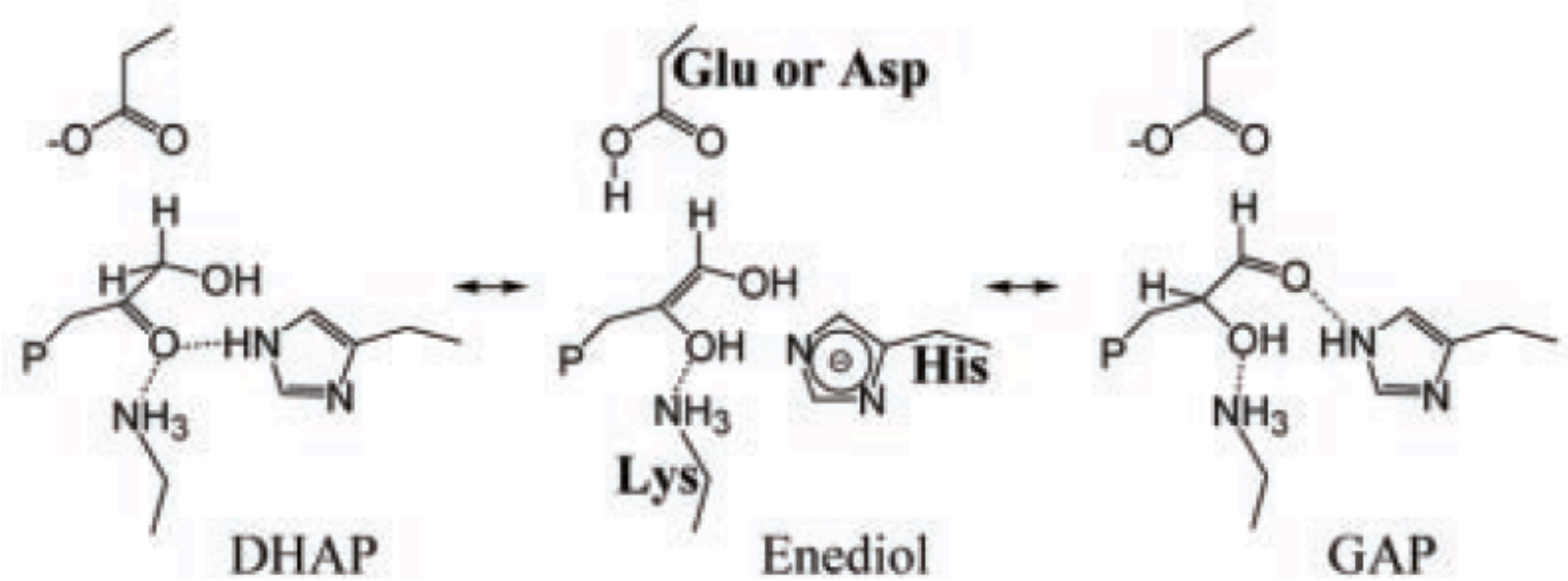


Redesigned enzyme cleaves the new recognition site ~10,000 times more effectively than wt I-MsoI homing endonuclease and with comparable target discrimination.

Ashworth, J et. al., *Nature* 441, 656-659 (2006).

Engineering Proteins by Design

Turning catalytically inert ribose-binding protein (RBP) into a triose phosphate isomerase (TIM)



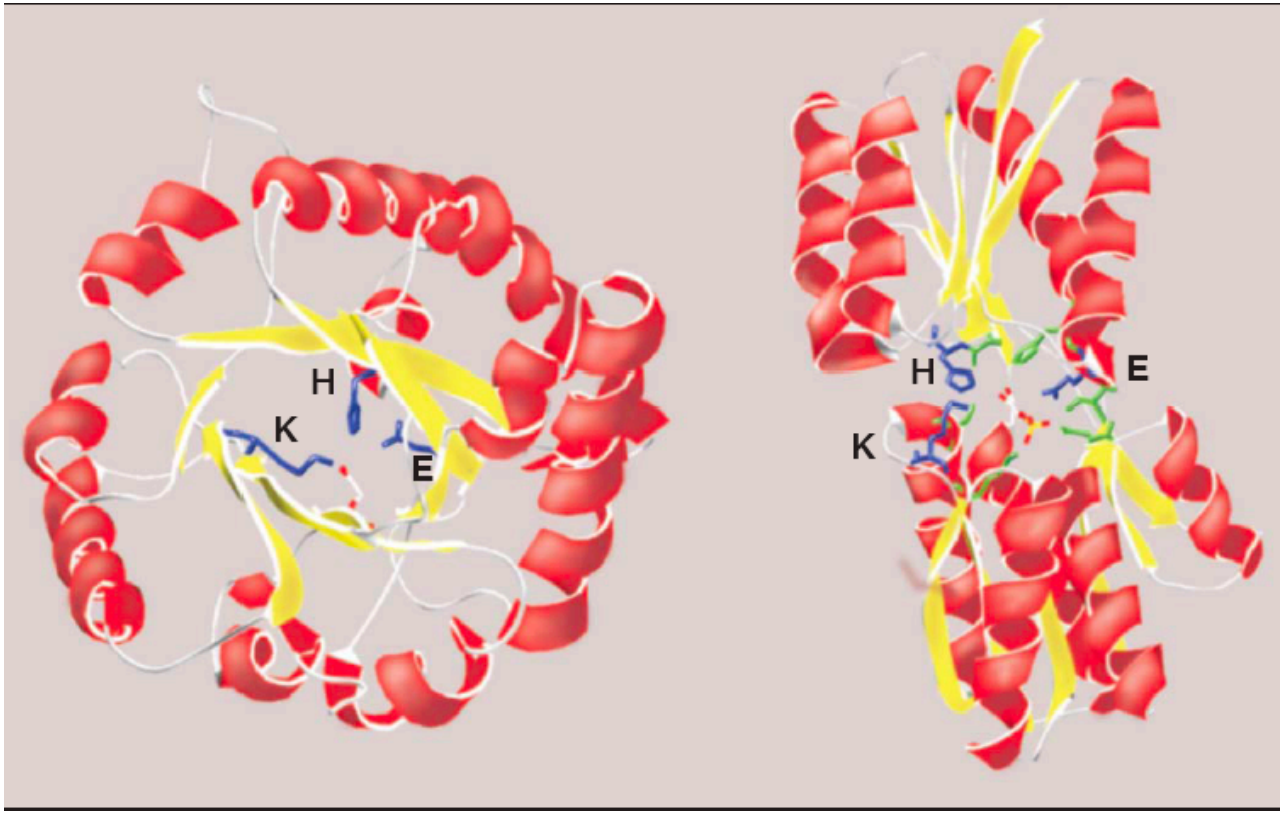
TIM mechanism is very sophisticated

Involves acid-based catalysis mediated by a catalytic triad (Glu, His, Lys)

The abstraction of the C1 proton is energetically highly unfavourable and the enediol can lose the phosphate by β-elimination. TIM avoids this side reaction by closing the active site with a mobile loop which acts as a lid.

Engineering Proteins by Design

Turning catalytically inert ribose-binding protein (RBP) into a triose phosphate isomerase (TIM)



Native TIM

NovoTIM1.2
(RBP containing a modeled TIM-like active site)

Required engineering of the catalytic triad (Glu, His, Lys) in precise orientation and optimisation of potential active site for the enediol intermediate

Dwyer et al, *Science* 304, 1967-1971 (2004)

Engineering Proteins by Design

Turning catalytically inert ribose-bindig protein (RBP) into a triose phosphate isomerase (TIM)

NovoTIM1.2 showed significant catalytic activity

	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	k_{cat}/k_{uncat}	wtk_{cat}/mutk_{cat}	wtk_{cat}/K_M/mutk_{cat}/K_M
Wild-type TIM	487	1600	3.0×10^5	2.4×10^9	4870	536
NovoTIM1.2	0.10	180	5.6×10^2	5.0×10^5		

But is still much less active than wt TIM

Dwyer et al, *Science* 304, 1967-1971 (2004)

Engineering Proteins by Design

Turning catalytically inert ribose-binding protein (RBP) into a triose phosphate isomerase (TIM) - all an artifact (or possibly fraud)

Retraction

WE WISH TO RETRACT OUR REPORT "COMPUTATIONAL DESIGN OF A BIOLOGICALLY ACTIVE enzyme" (*1*), which describes triose phosphate isomerase activity in a computationally redesigned ribose-binding protein (RBP) from *E. coli*. Dr. John P. Richard (Department of Chemistry, Department of Biochemistry, The State University of New York at Buffalo), to whom we provided clones encoding the novoTIM activity, has brought to our attention that the triose phosphate isomerase activity observed in our reported preparations can be attributed to a wild-type TIM impurity—seen in preparations that use a continuous rather than stepwise imidazole gradient (as in the original paper) or that add a second sepharose column. Richard's reanalysis has now also been confirmed by others in the Hellinga laboratory. The interpretations in the original report were based on lack of observed activity in mutant, engineered enzyme that bound substrate, but lacked catalytic residues. Variations in expression levels of designed proteins relative to the amount of contaminating endogenous protein might account for the pattern of observed activities that led to our erroneous conclusions. The in vivo experiments have not been reexamined.

We deeply regret that our report of a designed enzyme activity does not live up to closer scrutiny. Nevertheless, we remain optimistic that the problem of structure-based design of enzyme activity will be solved and that novel catalysts will be produced in conjunction with computationally based methods.

MARY A. DWYER,[1] LOREN L. LOOGER,[2] HOMME W. HELLINGA[3]

[1]Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, NC 27710, USA. [2]Howard Hughes Medical Institute, Janelia Farm, Chevy Chase, MD 20815–6789, USA. [3]Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA.

Reference

1. M. A. Dwyer, L. L. Looger, H. W. Hellinga, *Science* **304**, 1967 (2004).

K_M of NovoTIM1.2 found to be identical to wild-type TIM

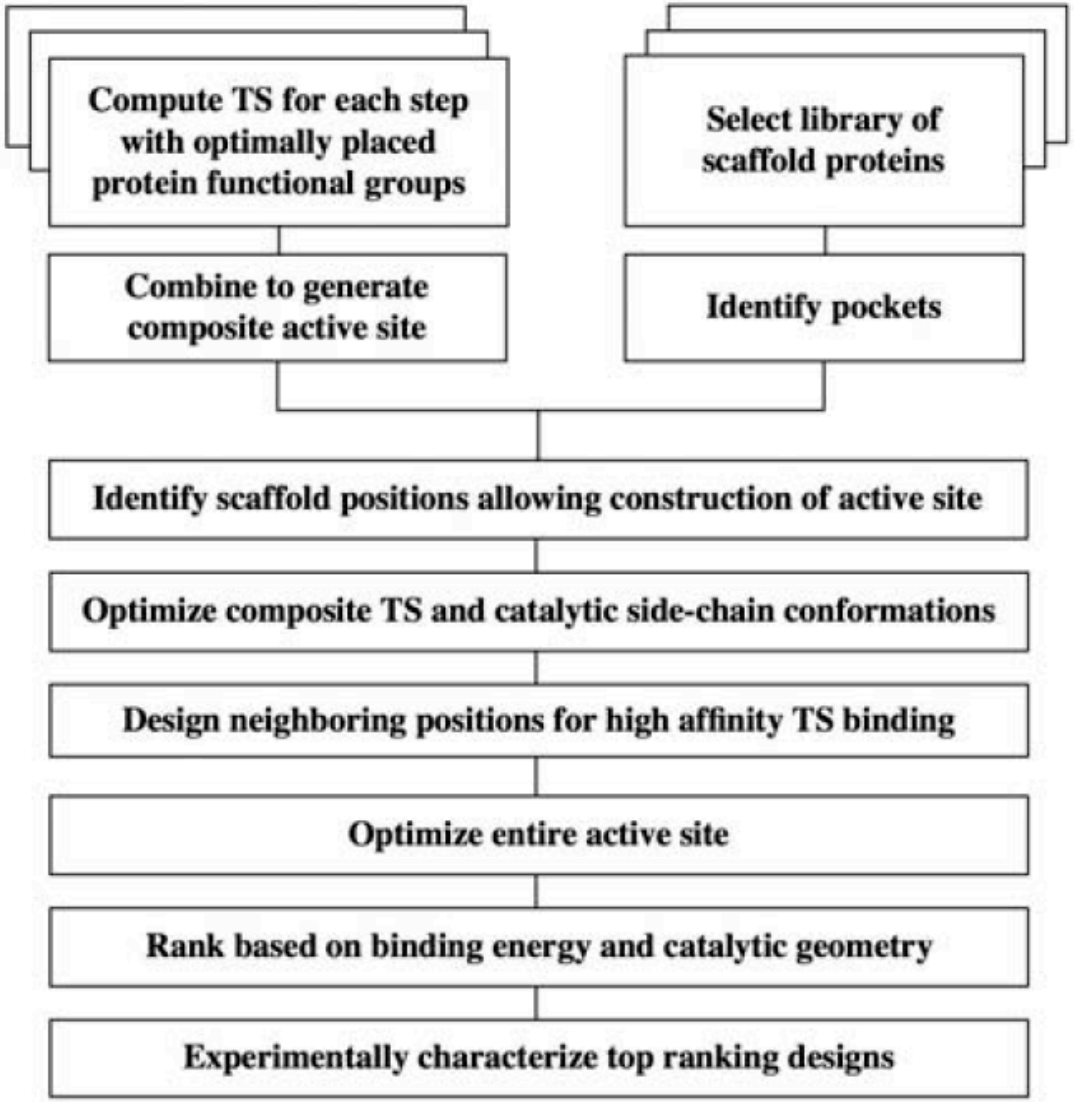
The activity seen was simply wild-type TIM impurity

For more information see: http://www.nature.com/nature/journal/v453/n7193/full/453258b.html

http://www.nature.com/news/2008/080514/full/453275a.html

Engineering Proteins by Design

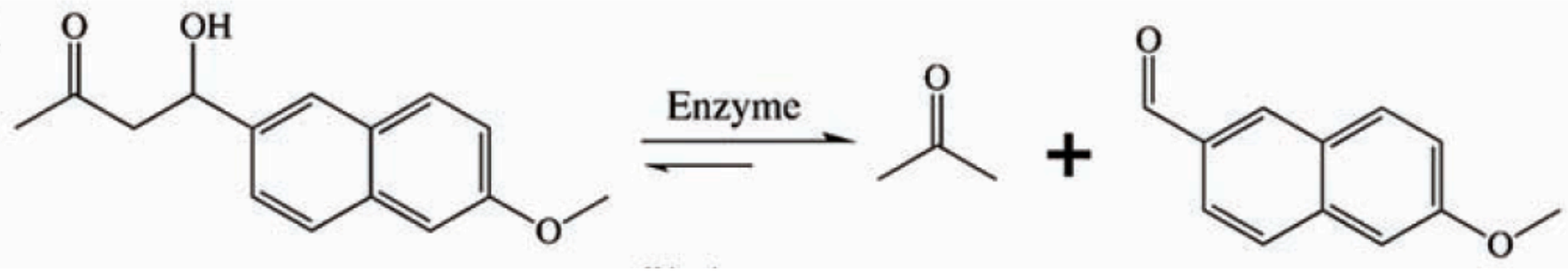
De-novo computational design of retro-aldol enzymes



Jiang, L. et al., Science 319, 1387-1391 (2008)

Engineering Proteins by Design

De-novo computational design of retro-aldol enzymes



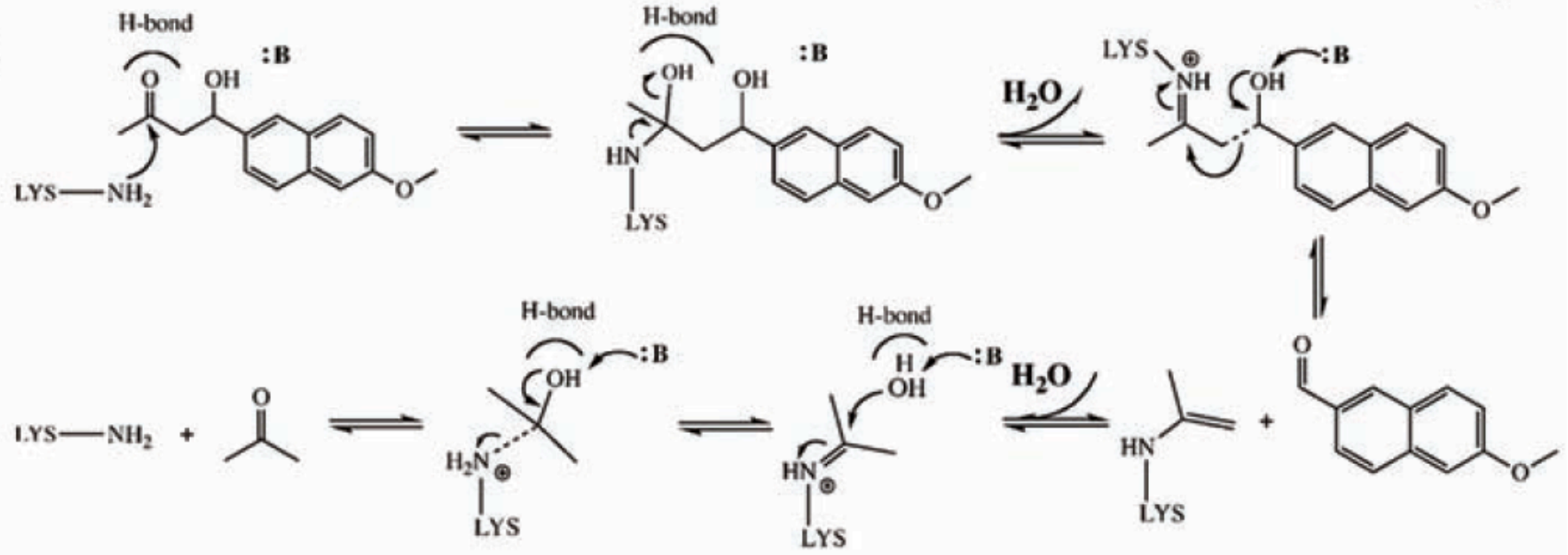
The retro-aldol reaction

The substrate was non-natural - not used by natural enzymes
This avoids contamination problems of the sort seen with NovoTIM1.2

Jiang, L. et al., Science 319, 1387-1391 (2008)

Engineering Proteins by Design

De-novo computational design of retro-aldol enzymes

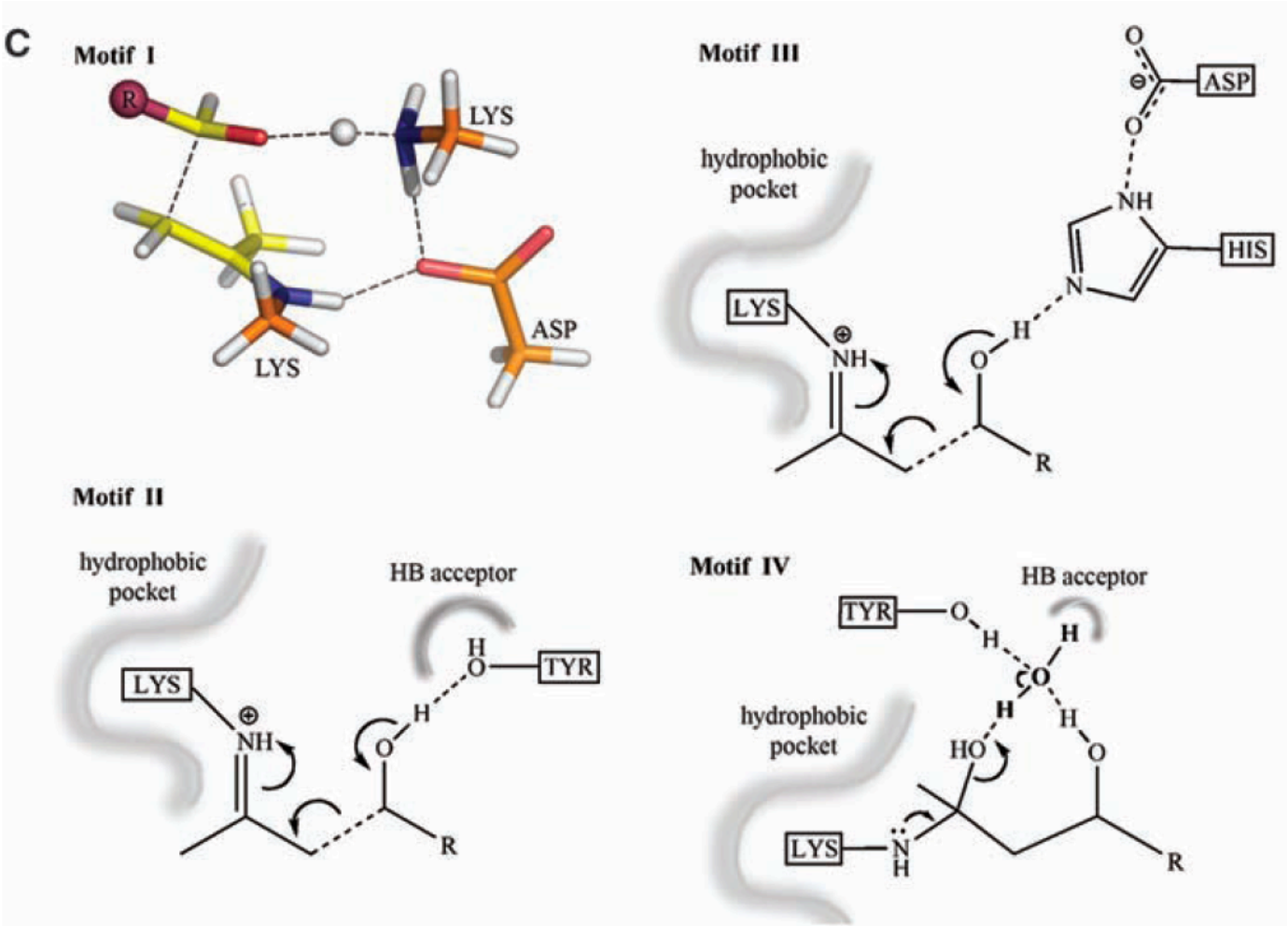


General description of the aldol reaction pathway with a nucleophilic lysine and general acid base chemistry

Jiang, L. et al., Science 319, 1387-1391 (2008)

Engineering Proteins by Design

De-novo computational design of retro-aldol enzymes



Active-site motifs with quantum mechanical optimised structures

Jiang, L. et al., Science 319, 1387-1391 (2008)

Engineering Proteins by Design

De-novo computational design of retro-aldol enzymes

Motif	Catalytic lysine environment	Carbinolamine stabilization	Proton abstraction	Number tested	Number forming enaminone	Number of active designs	Rate enhancement
I	Polar	NC	Lys-Asp dyad	12	2	0	<4
II	Hydrophobic	NC	Tyr	9	1	0	<4
III	Hydrophobic	H-bond acceptor/donor	His-Asp dyad	13	10	10	10^2~10^3
IV	Hydrophobic	Water, H-bond acceptor	Water	38	20	22	10^3~10^4

Enzyme activity for different active-site motifs

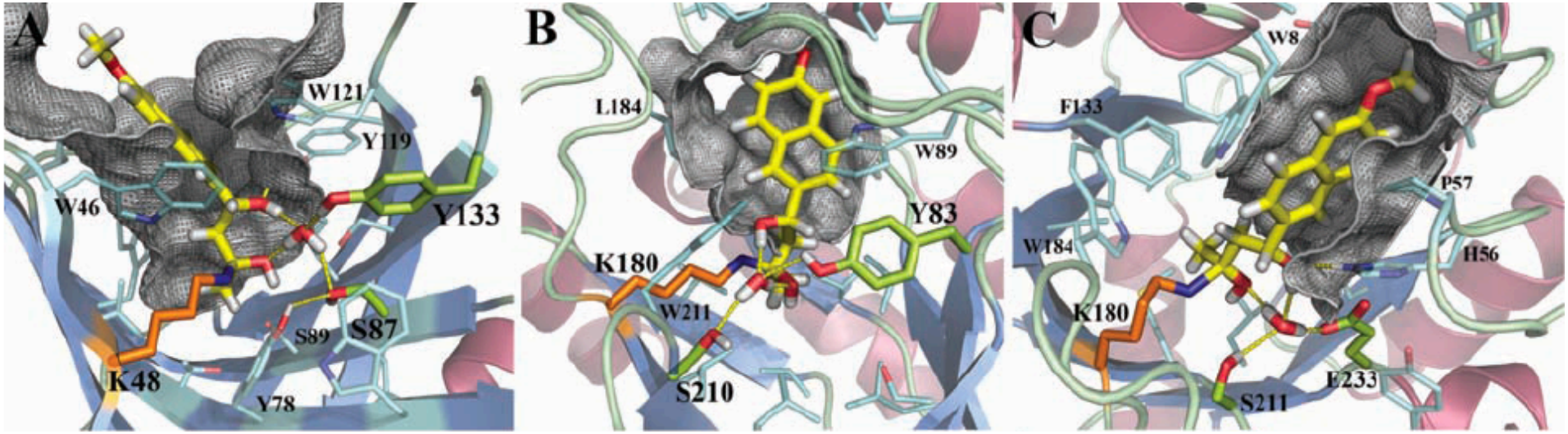
72 designs tested using 4 different catalytic motifs
32 showed detectable activity
Active variants contained: 4 different catalytic motifs
10 different protein folds

The designs using H_2O molecules to shuffle protons were the most successful (better than charged side-chain networks)

Jiang, L. et al., Science 319, 1387-1391 (2008)

Engineering Proteins by Design

De-novo computational design of retro-aldol enzymes



RA60
Catalytic motif IV
Jelly roll scaffold

RA46
Catalytic motif IV
TIM-barrel scaffold

RA45
Catalytic motif IV
TIM-barrel scaffold

Examples of design models for active designs

Groups important for catalysis are highlighted:

Nucleophilic imine-forming lysine = orange

Transition state (TS) = yellow

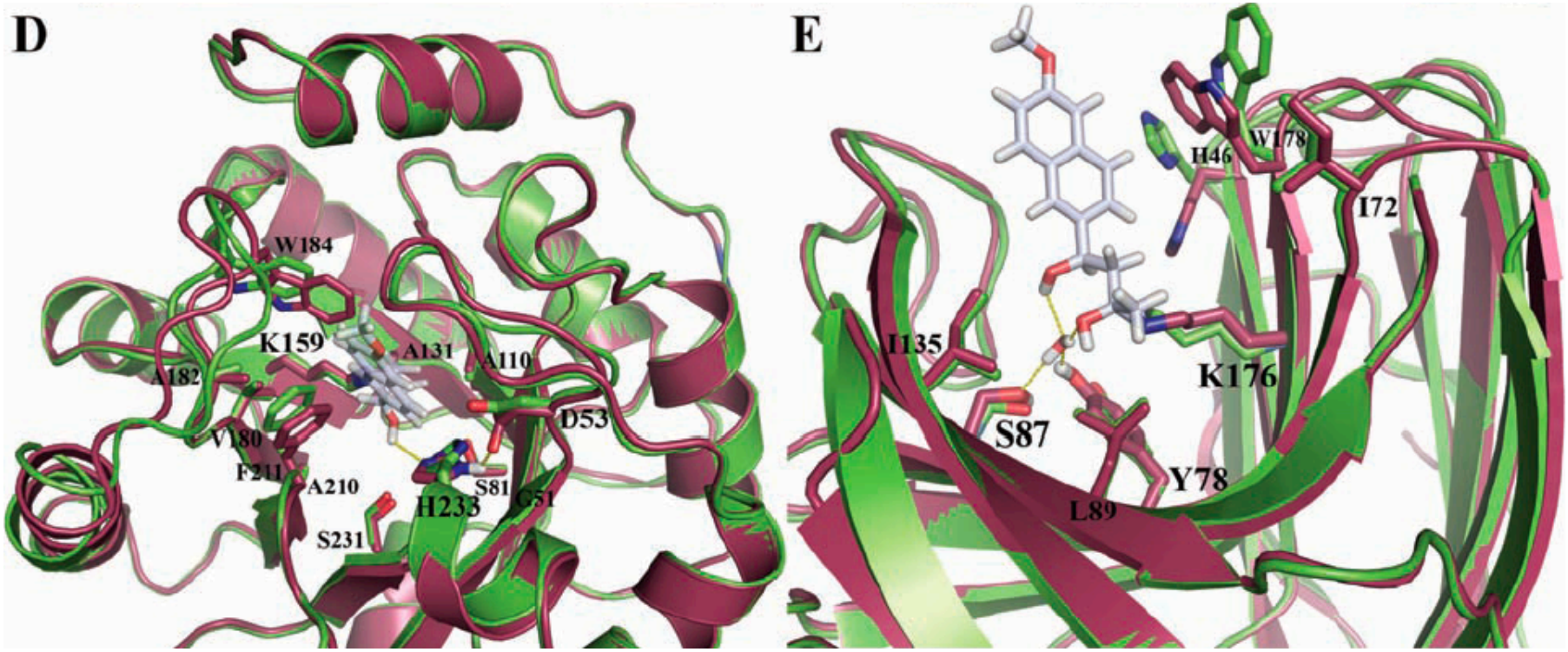
H-bonding groups = light green

The catalytic water molecule is also shown

Jiang, L. et al., Science 319, 1387-1391 (2008)

Engineering Proteins by Design

De-novo computational design of retro-aldol enzymes



RA22
Catalytic motif III Cα RMSD 0.62Å
TIM-barrel scaffold

RA61
Catalytic motif IV Cα RMSD 0.42Å
Jelly roll scaffold

Overlay of design models on x-ray structures
Quite good fit

Jiang, L. et al., Science 319, 1387-1391 (2008)

Engineering Proteins by Design

De-novo computational design of retro-aldol enzymes

Design	k_{cat} ($\times 10^{-3}$ min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat}/k_{uncat}*
RA22	3.1 ± 0.3 (b)	480 ± 130 (b)	0.11 ± 0.03 (b)	8.1×10^3 (b)
	0.5 ± 0.1 (s)	450 ± 210 (s)	0.018 ± 0.006 (s)	1.2×10^3 (s)
RA34	4.2 ± 1.1 (b)	620 ± 180 (b)	0.11 ± 0.01 (b)	1.1×10^4 (b)
	0.6 ± 0.1 (s)	600 ± 140 (s)	0.016 ± 0.004 (s)	1.5×10^3 (s)
RA45	2.3 ± 0.2	430 ± 48	0.091 ± 0.004	6.0×10^3
RA46	0.62 ± 0.5	290 ± 60	0.037 ± 0.006	1.6×10^3
RA60	9.3 ± 0.9	510 ± 33	0.30 ± 0.06	2.4×10^4
RA61	9.0 ± 1.0	210 ± 50	0.74 ± 0.11	2.3×10^4

*$k_{uncat} = 3.9 \times 10^{-7}$ min^{-1} (*25*).

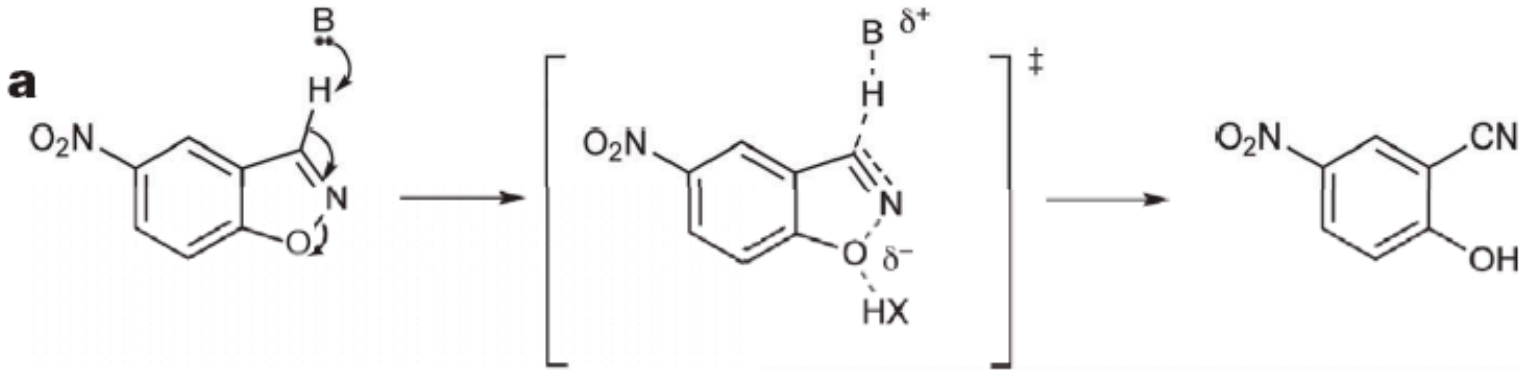
Kinetic parameters of some of the best designs.
b, burst phase; s, steady state

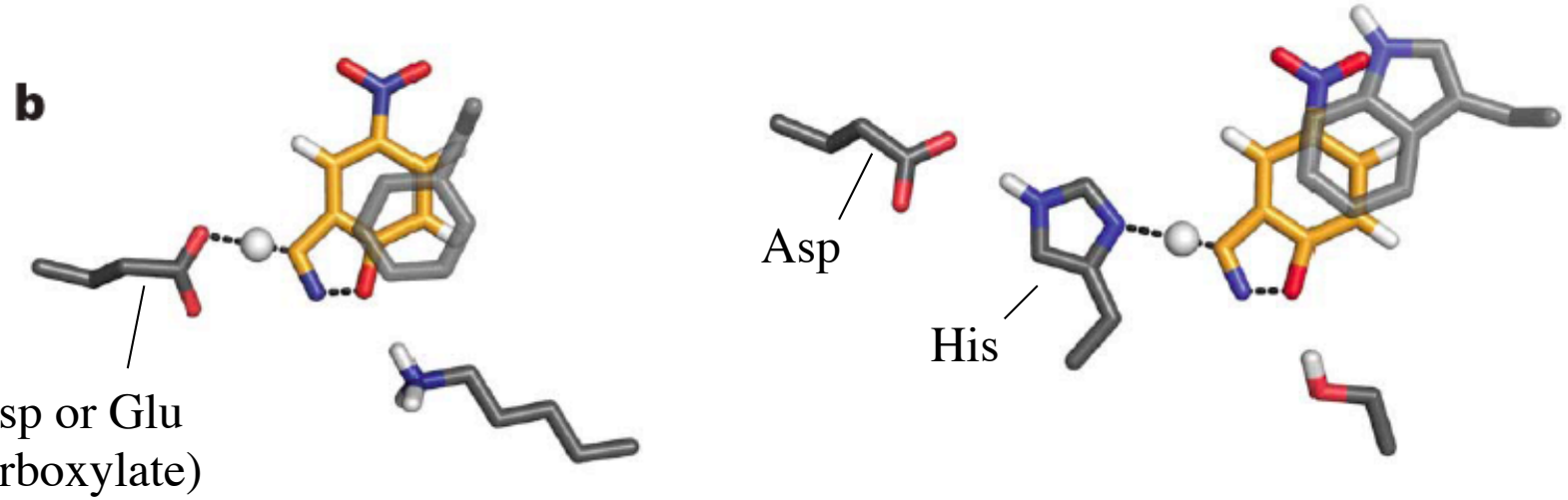
Still very slow compared to natural enzymes

Jiang, L. et al., Science 319, 1387-1391 (2008)

Engineering Proteins by Design

Kemp elimination





Two choices for catalytic base

- A carboxylate (left)
- A His-Asp (or Glu) dyad (right)

Röthlisberger, D. et al., Nature 453, 190-195 (2008)

Engineering Proteins by Design

Kemp elimination

Table 1 | Kinetic parameters of designed enzymes

Design	PDB code of template	Base	Hydrogen-bond donor	π-stack	k_{cat} (s^{-1}; mean ± s.d.)	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/k_{uncat}*
KE07	1thf	E101	K222	W50	0.018 ± 0.001	1.4 ± 0.1	12.2 ± 0.1	1.6×10^4
KE07	1thf	E101A	K222	W50	0.0009 ± 0.0004	0.29 ± 0.14	3.5 ± 1.0	7.8×10^2
KE07	1thf	E101	K222A	W50	0.030 ± 0.004	1.3 ± 0.2	22.7 ± 2.7	2.6×10^4
KE10	1a53	E178	None	W210	NA	NA	51.6 ± 4.0	NA
KE10	1a53	E178Q	None	W210	NA	NA	0.17 ± 0.17	NA
KE15	1thf	D48	None	Y126	0.022 ± 0.003	0.63 ± 0.09	35.1 ± 4.8	1.9×10^4
KE15	1thf	D48A	None	Y126	NA	NA	0.05 ± 0.07	NA
KE16	1thf	D48	K201	Y126	0.006 ± 0.001	4.2 ± 0.1	13.4 ± 1.4	5.2×10^3
KE16	1thf	D48A	K201	Y126	ND	ND	ND	ND
KE59	1a53	E231	None	W110	0.29 ± 0.11	1.8 ± 0.6	163 ± 21	2.5×10^5
KE59	1a53	E231Q	None	W110	NA	NA	0.003 ± 0.002	NA
KE59	1a53	E231	G131S	W110	NA	NA	17.9 ± 0.7	NA
KE61	1h61	E100	None	W184	NA	NA	7.9 ± 0.6	NA
KE61	1h61	E100A	None	W184	NA	NA	0.7 ± 0.1	NA
KE70	1jcl	H16–D44	S137	Y47	0.16 ± 0.05	2.1 ± 0.8	78.3 ± 13.7	1.4×10^5
KE70	1jcl	H16A–D44	S137	Y47	NA	NA	0.02 ± 0.03	NA
KE70	1jcl	H16–D44N	S137	Y47	0.06 ± 0.03	2.3 ± 1.5	29.8 ± 4.6	5.2×10^4
KE71	1a53	H89–D85	Y210	W184	NA	NA	5.9 ± 0.2	NA
KE71	1a53	H89A–D85	Y210	W184	NA	NA	0.4 ± 0.2	NA
KE71	1a53	H89–D85N	Y210	W184	NA	NA	1.0 ± 0.2	NA

* k_{uncat} (1.16×10^{-6} s^{-1}) was determined in HEPES buffer at pH 7.25, and extrapolated to zero buffer concentration. The means and standard deviations of the kinetic parameters were calculated from at least three independent measurements. NA, not applicable; ND, not determined.

59 designs in 17 scaffolds tested - 8 active

Still very slow compared to natural enzymes

Röthlisberger, D. et al., Nature 453, 190-195 (2008)

Directed Evolution

Directed Evolution

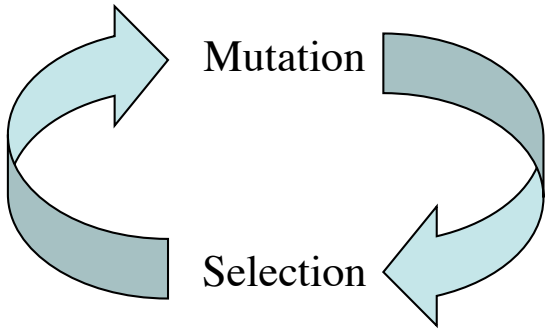
Evolution is a very powerful process

Natural enzymes are remarkable catalysts

Examples of catalytic efficiencies of natural enzymes.

#	Enzyme	Uncatalysed half-life $t_{1/2}^{uncat}$	Turnover number k_{cat} (s^{-1})	Catalysed half-life $t_{1/2}^{cat}$ (ms)	Specificity factor k_{cat}/K_M (s^{-1} M^{-1})	Rate acceleration k_{cat}/k_{uncat}
1	OMP decarboxylase*	7.8×10^7 years	39	18	5.6×10^7	1.4×10^{17}
2	Acetylcholine esterase†	~3 years	$>10^4$	<0.07	$>10^8$	$\sim 10^{13}$
3	Triosephosphate isomerase*	1.9 days	4300	0.16	2.4×10^8	1.0×10^9
4	Chorismate mutase*	7.4 hours	50	13.8	1.1×10^6	1.9×10^6
5	*Tetrahymena* group I ribozyme‡	~430 years	~5.8	120	1.5×10^6	$\sim 10^{11}$

They have not been generated by design, but by Evolution



Repeated rounds of mutation and selection

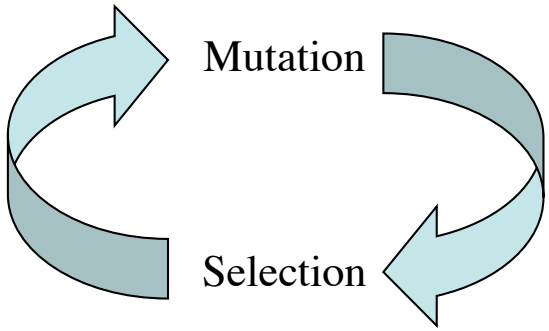
Directed Evolution

Evolution is a very powerful process

Frances Arnold

Nobel Prize in
Chemistry 2018
for
Directed Evolution

They have not been generated by design, but by Evolution

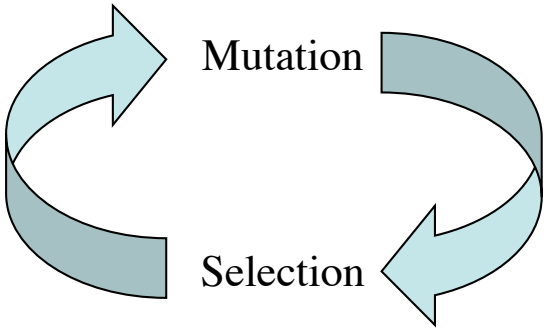


Repeated rounds of mutation and selection

Directed Evolution

Applying Darwinian evolutionary principles in the laboratory

Using evolutionary systems to evolve proteins (and nucleic acids) in the laboratory is called "Directed Evolution"



Directed evolution requires:

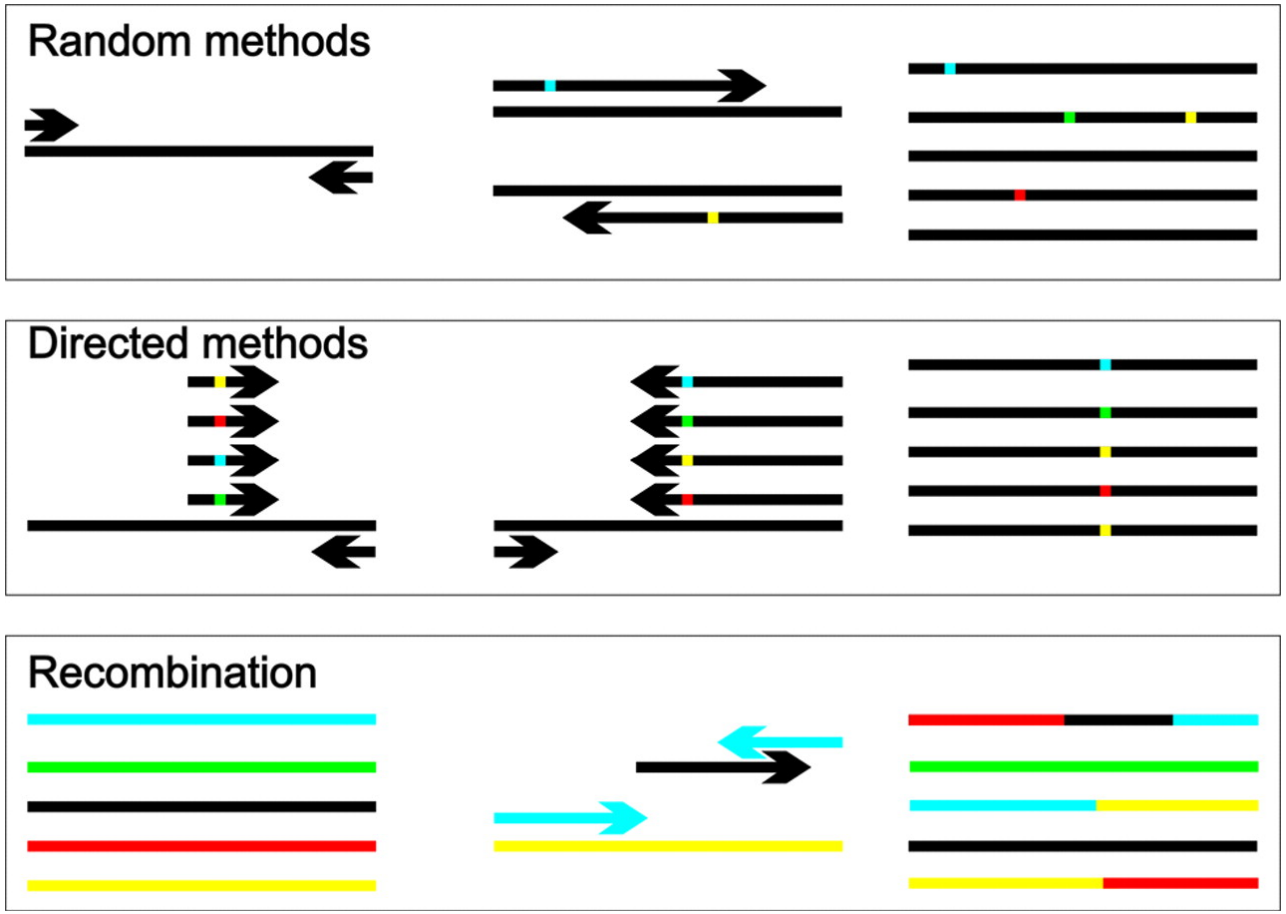
1. A system for generating genetic diversity - i.e. for mutation and recombination

2. A system to select for the desired activity - typically a binding or catalytic activity

Generating Genetic Diversity

Methods for the randomisation of DNA sequences

There are fundamentally three different strategies:



Generating Genetic Diversity

Classical strategies for generating diversity throughout a DNA sequence

"Classical" strategies - used for many years:

- UV irradiation
- Alkylating agents

Damage DNA which is incorrectly repaired leading to mutations

Mutator strains e.g. XL1-Red

Bacteria with one or more DNA repair pathways are defective leading to higher mutation rates

Advantages:

Simple - just passage DNA through strain.

Dissadvantages:

Low mutation rate - constructing a library with 1 or 2 mutations per gene may require multiple passages through the mutator strain

Hence, an alternative technique, error-prone PCR, is used in most directed evolution experiments

Generating Genetic Diversity

Generating diversity throughout a DNA sequence - error-prone PCR

Error-prone PCR

Based on increasing the error rate of Taq DNA polymerase

1. Introduce small amount of Mn^{2+} (in place of Mg^{2+}) and include biased concentrations of dNTPs - over-representation of dGTP and dTTP
    - Gives error rates of ~1nt/kb

2. For higher rates of mutagenesis can use nucleoside triphosphate analogues 8-oxo-dGTP and dPTP
    - Gives error rates of up to ~200nt/kb

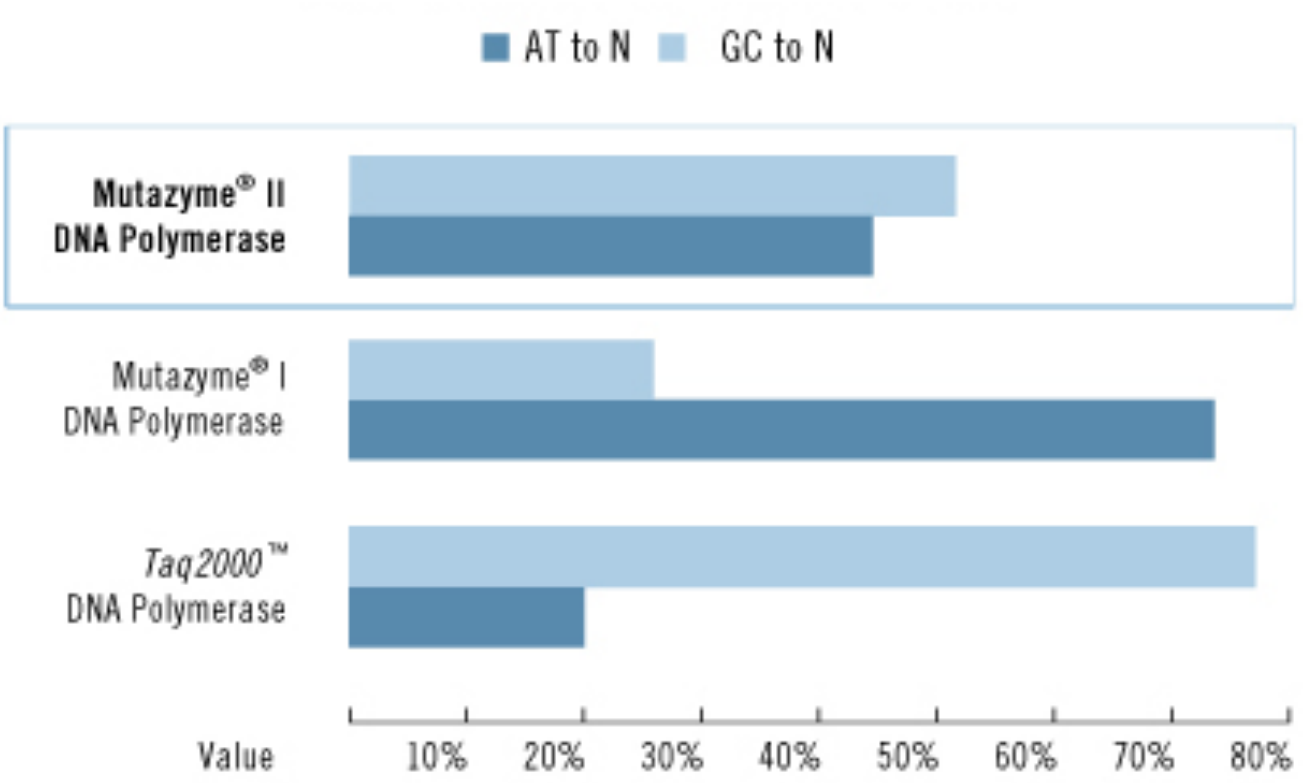
3. Error-prone DNA polymerase - e.g. Mutazyme II (Stratagene)
    - Gives error rates of ~ 1 to 16 bases per kb

Generating Genetic Diversity

Generating diversity throughout a DNA sequence - error-prone PCR

Three sources of bias in error-prone PCR libraries

1. Bias in misincorporation by the polymerase - solutions:
    - Use combination of techniques e.g. Taq and Mutazyme I
    - Use new less bias prone enzymes e.g. Mutazyme II



Generating Genetic Diversity

Generating diversity throughout a DNA sequence - error-prone PCR

2. Codon bias - due to the nature of the genetic code.
   - Single point mutations lead to a bias in the variant amino-acids coded

Table of Standard Genetic Code

	T	C	A	G
T	TTT Phe (F) TTC Phe (F) TTA Leu (L) TTG Leu (L)	TCT Ser (S) TCC Ser (S) TCA Ser (S) TCG Ser (S)	TAT Tyr (Y) TAC TAA STOP TAG STOP	TGT Cys (C) TGC TGA STOP TGG Trp (W)
C	CTT Leu (L) CTC Leu (L) CTA Leu (L) CTG Leu (L)	CCT Pro (P) CCC Pro (P) CCA Pro (P) CCG Pro (P)	CAT His (H) CAC His (H) CAA Gln (Q) CAG Gln (Q)	CGT Arg (R) CGC Arg (R) CGA Arg (R) CGG Arg (R)
A	ATT Ile (I) ATC Ile (I) ATA Ile (I) ATG Met (M) START	ACT Thr (T) ACC Thr (T) ACA Thr (T) ACG Thr (T)	AAT Asn (N) AAC Asn (N) AAA Lys (K) AAG Lys (K)	AGT Ser (S) AGC Ser (S) AGA Arg (R) AGG Arg (R)
G	GTT Val (V) GTC Val (V) GTA Val (V) GTG Val (V)	GCT Ala (A) GCC Ala (A) GCA Ala (A) GCG Ala (A)	GAT Asp (D) GAC Asp (D) GAA Glu (E) GAG Glu (E)	GGT Gly (G) GGC Gly (G) GGA Gly (G) GGG Gly (G)

For example, in Valine codon:

1 point mutation	➡ can only encode F, L, I, A, D, G	(6)
2 point mutations	➡ can also encode C, S, P, H, R, N, T, M, E, Y	(16)
3 point mutations	➡ can also encode Q, W, K	(19)

Generating Genetic Diversity

Generating diversity throughout a DNA sequence - error-prone PCR

3. Amplification bias

Mutations which arise in early cycles of PCR amplification will be over-represented

An extreme case - a PCR starting from a single DNA molecule:
If a mutation occurs in the first step it would be present in 25% of product molecules

Partial solutions:

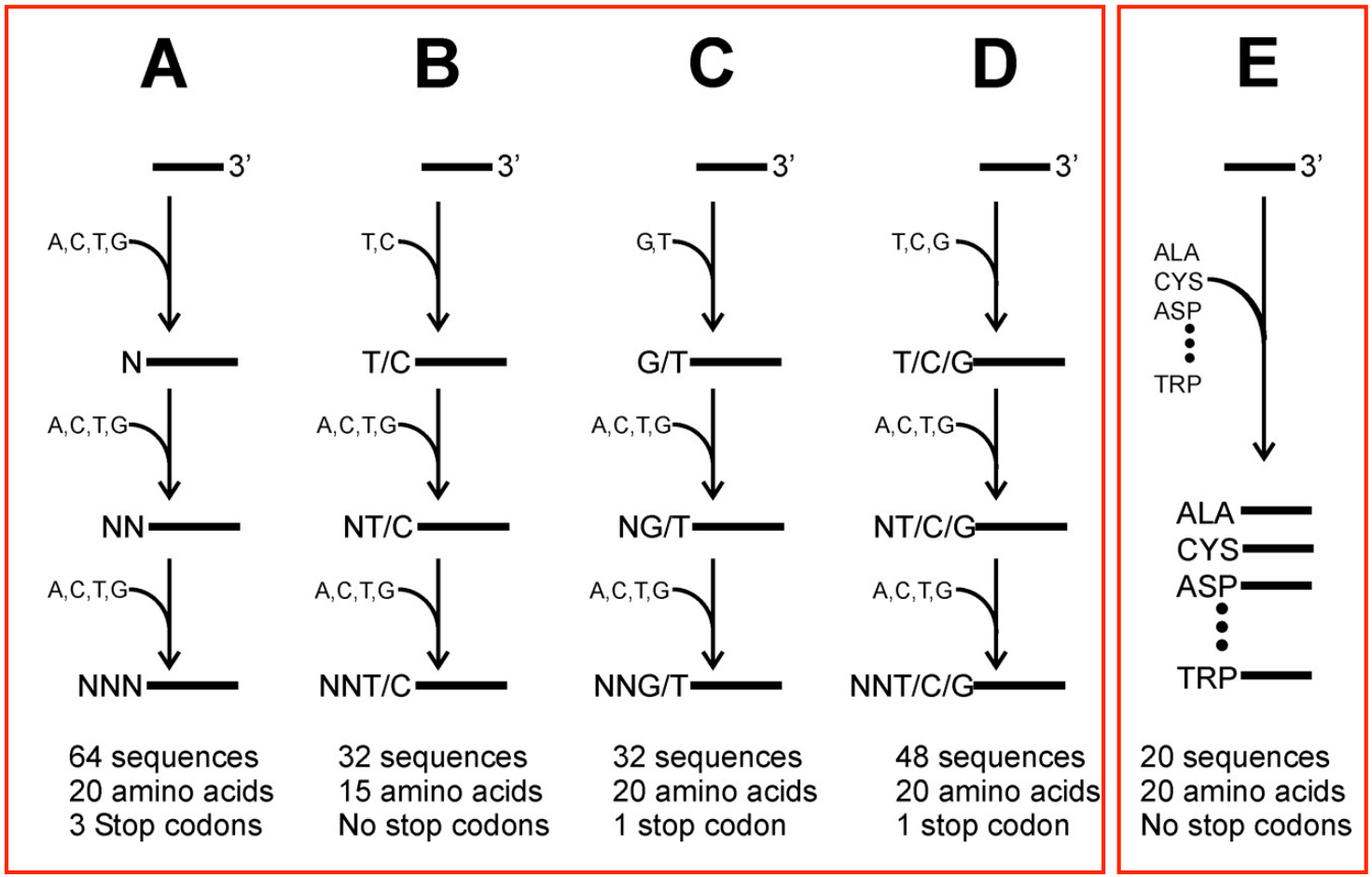
- Start the error-prone PCR from many DNA molecules
- Perform multiple independent error-prone PCRs
- Reduce number of amplification cycles (unfortunately the number of cycles is also one of the best ways to control the level of mutagenesis)

Generating Genetic Diversity

Directed diversity: oligonucleotide-based methods

Saturation mutagenesis of one or more codons

- One can synthesise DNA oligonucleotides with any chosen sequence giving great flexibility
- There are several approaches to randomising synthetic DNA



Mixed nucleotides

trinucleotide phosphoramidites

Generating Genetic Diversity

Directed diversity: oligonucleotide-based methods

"Doping"

One can choose to have locations where the codons are mostly wild-type and only mutated occasionally.

e.g. using an NNN strategy to randomise the glycine codon GGG, for each position use:

- Saturation mutagenesis - - use 25%A, 25%C, 25%T, 25%G
- Doping - use 5%A, 5%C, 5%T, 85%G

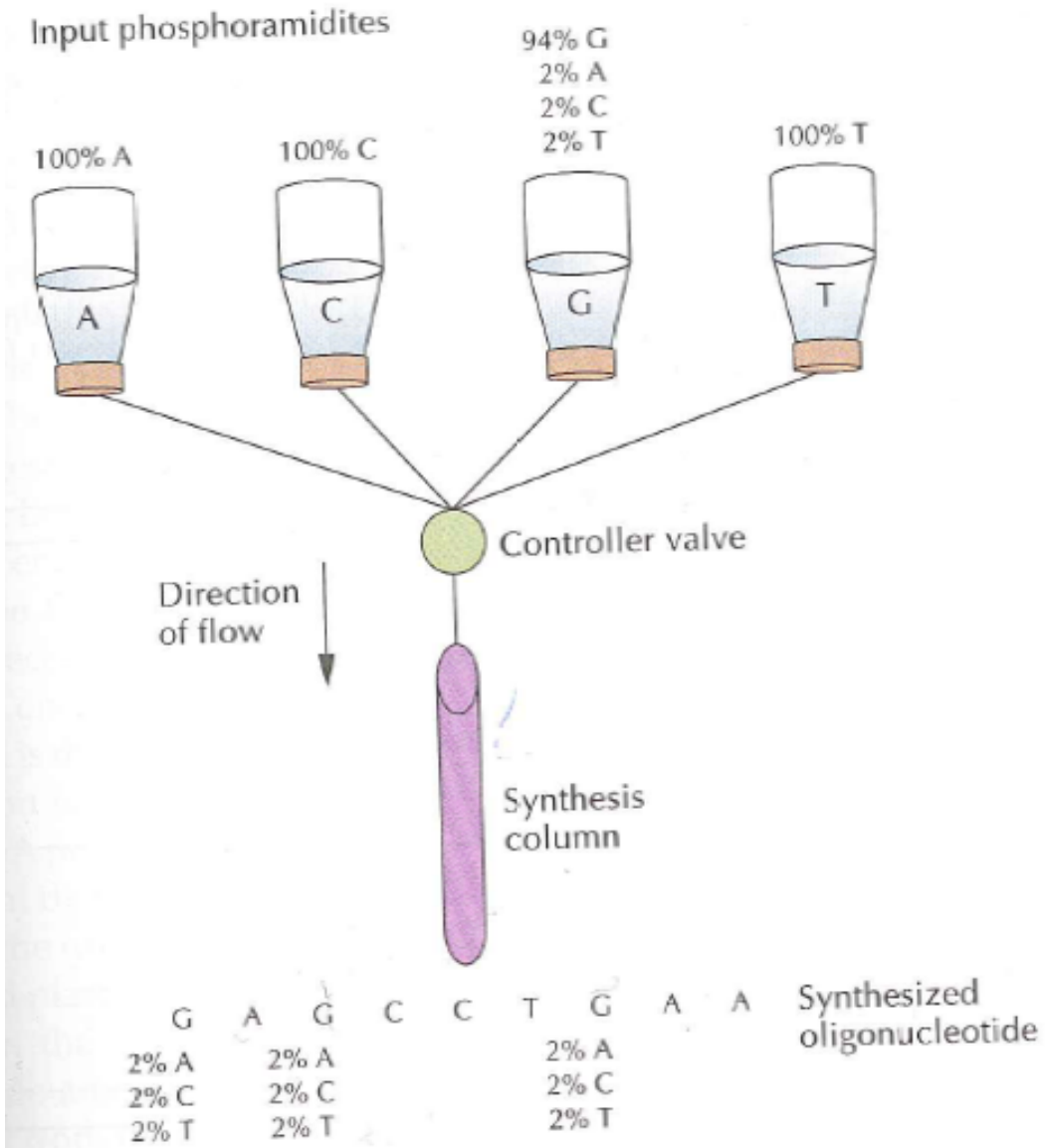
The percentage of wild-type nucleotide can be varied depending on the level of mutagenesis required.

This strategy results in codon bias (as with error-prone PCR) - except when using trinucleotide phosphoramidites.

The codon bias can be changed deliberately to allow (or favour) incorporation of certain amino-acids

Generating Genetic Diversity

Mutagenesis with doped oligonucleotide primers



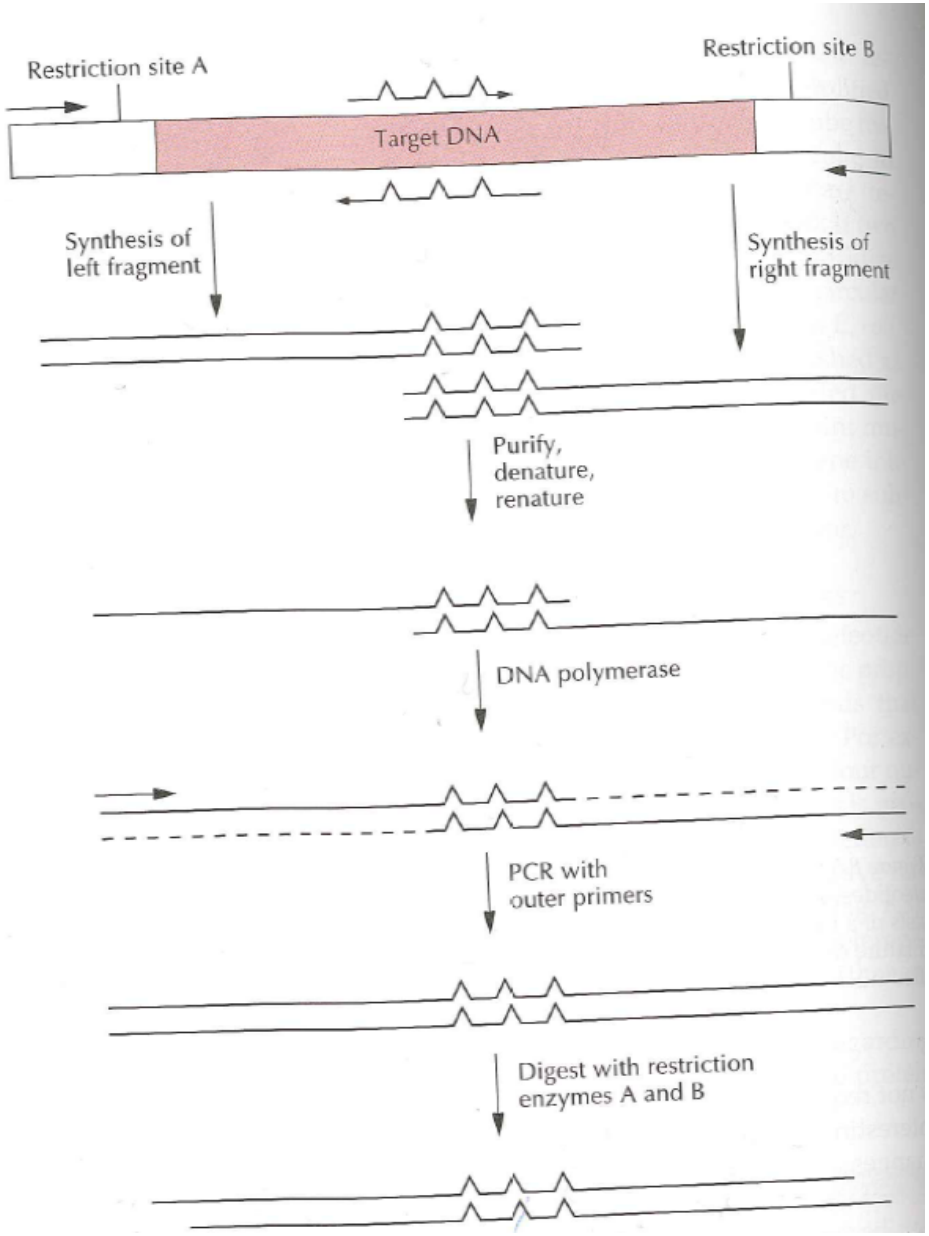
In this example, the flask with G phosphoramidite contains a mixture of:

94% G
2% A
2% G
2% T

Thus, at positions which are normally G, the incorporated nucleotide is still mostly G, but occasionally A, C or T.

Generating Genetic Diversity

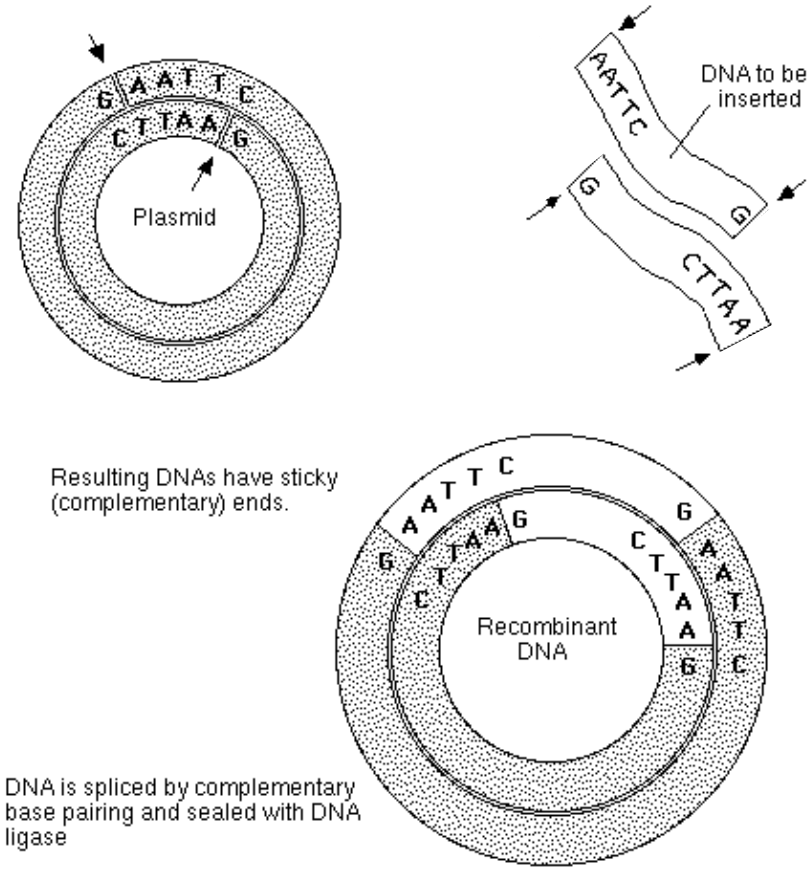
Mutagenesis of a target DNA using degenerate oligonucleotides and PCR



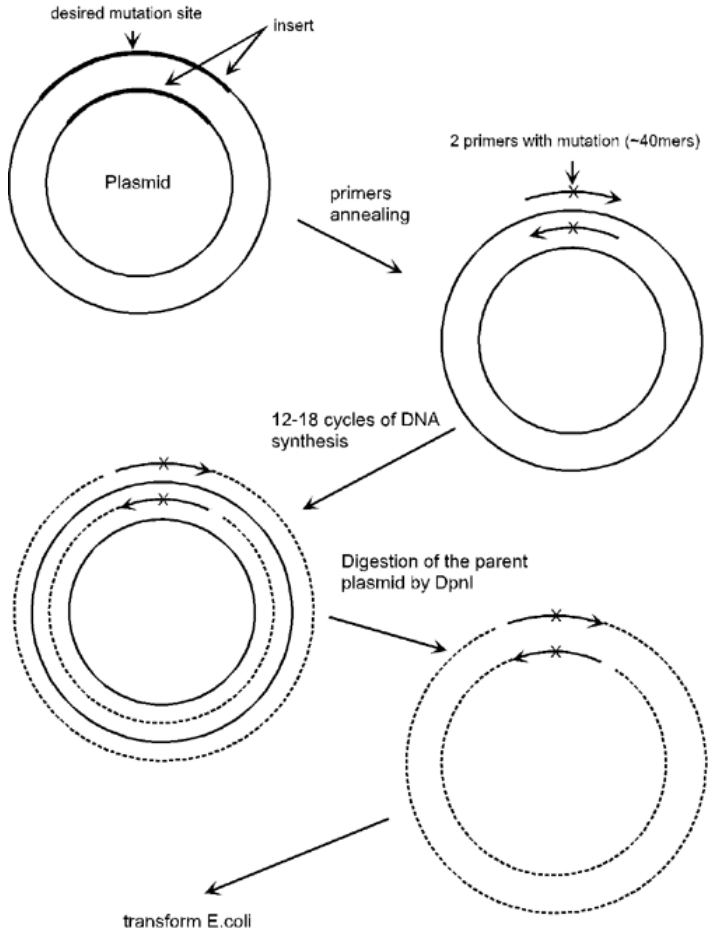
Generating Genetic Diversity

Incorporating synthetic DNA in full length DNA

2 basic strategies



Ligation



PCR-based
e.g. Quikchange

Recombination of DNA Sequences

The advantage of sex and recombination

In nature, sexual lifestyles, which at the genetic level are characterised by DNA recombination, are far more common than asexual ones.

This is despite some inherent detriments of sex in species with differentiated sexes, including:

- the twofold 'cost of producing males'- due to the reduction in the growth rate of a sexual population when males do not provide resources that increase the fecundity of their mates

- the twofold 'cost of meiosis'- which reduces parent–offspring relatedness from 1, in a female that reproduces parthanogenetically, to 0.5 in a sexually reproducing female

However, the driving mechanism for sex and recombination is still hotly debated - mathematical modelling indicates that sex should only be advantageous under certain rather special conditions

Recombination of DNA Sequences

Why all the sex?

Numerous theories have been proposed to account for the near-ubiquity of sex, including:

- Allowing two beneficial mutations arising in different lineages to become united in a single genome which is fitter than either of the parental strains (the Fisher-Muller hypothesis).

- Purging deleterious mutations from a genome; without such a mechanism, a population might accumulate deleterious mutations faster than they can be purged by selection (H. J. Muller).

- A beneficial mutation would stand a better chance of surviving in a recombining than a non-recombining system since, in an asexual population, the fate of a mutation strongly depends on the genotype in which it arises, whereas in a sexual population it can be selected in many different genetic backgrounds; in the absence of recombination, beneficial genes may become trapped in a less-fit genome background (so called "background trapping") (R.A. Fisher).

- sex might reduce competition between beneficial alleles ("clonal interference") by uniting them to form more advantageous genotypes.

- Sex might accelerate adaptation to new environments by increasing genetic variation.

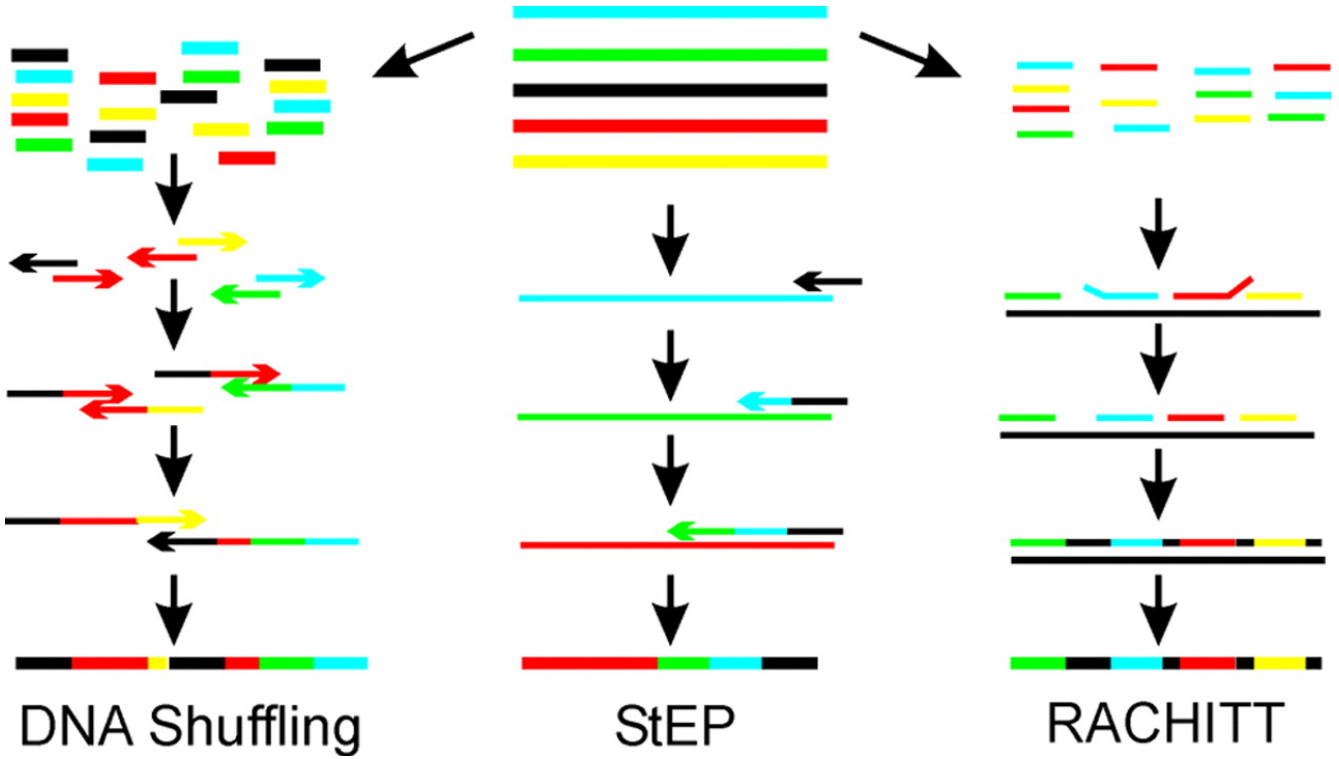
Recombination of DNA Sequences

Techniques for recombining homologous DNA sequences

Until 1993 there were no random recombination methods available for directed evolution.

The invention of DNA shuffling (or "Sexual PCR") changes this.

There are now several homology-based techniques for recombining DNA sequences:



They are all designed to create libraries of chimeric genes which are mosaics made up of fragments of genes with a fairly high degree of sequence homology.

Recombination of DNA Sequences

DNA shuffling - or "Sexual PCR"

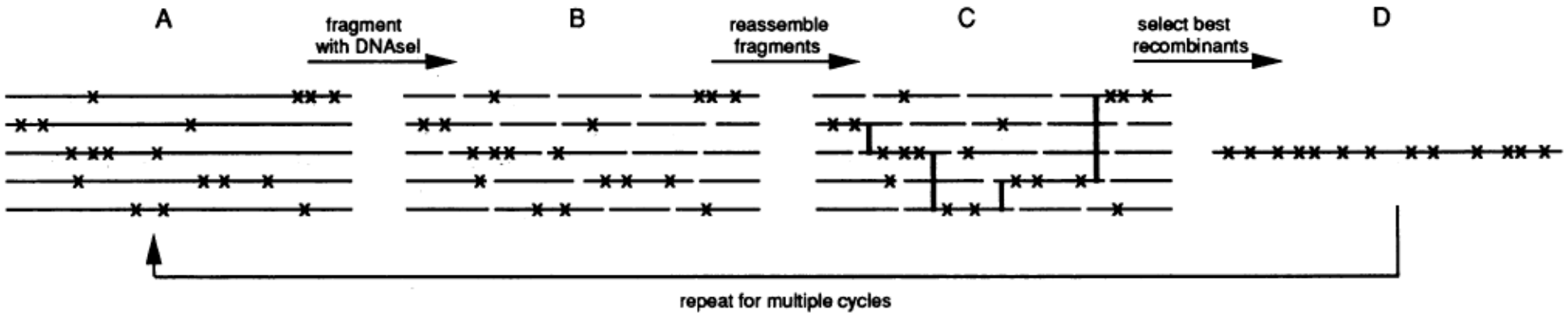
DNA shuffling (or "Sexual PCR") was the first - and is still the most widely used recombination technique.

It is very straightforward:

1. Fragment DNA from different genes with DNAseI

2. Repeat cycles of melting, annealing and extension to reassemble fragments - this works because one fragment can act as a primer on another due to short sequence complementarities (it is like a PCR with no oligonucleotide primers)

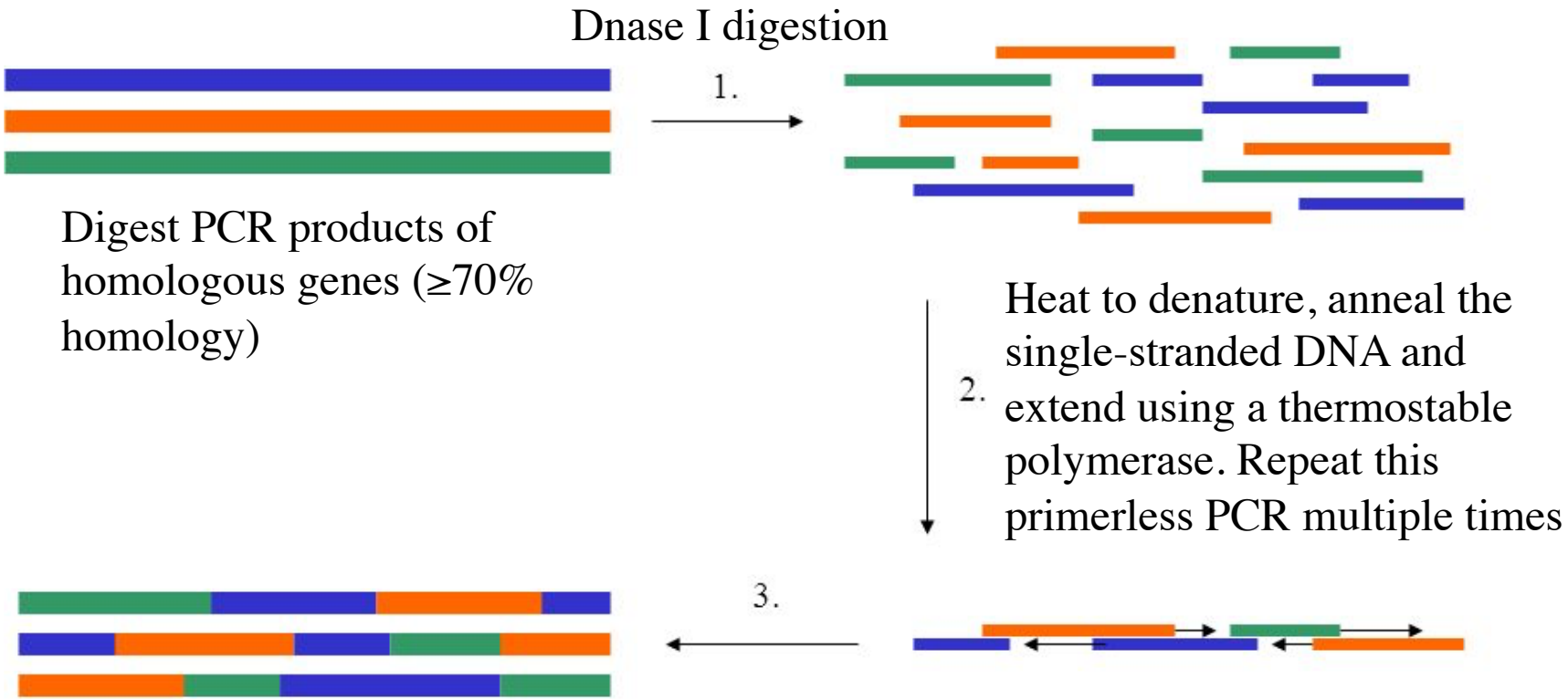
3. Amplify reassembled genes by PCR using oligonucleotide primers



Stemmer, W.P.C. *Proc. Natl. Acad. Sci USA* 91, 10747-10751 (1993).

Recombination of DNA Sequences

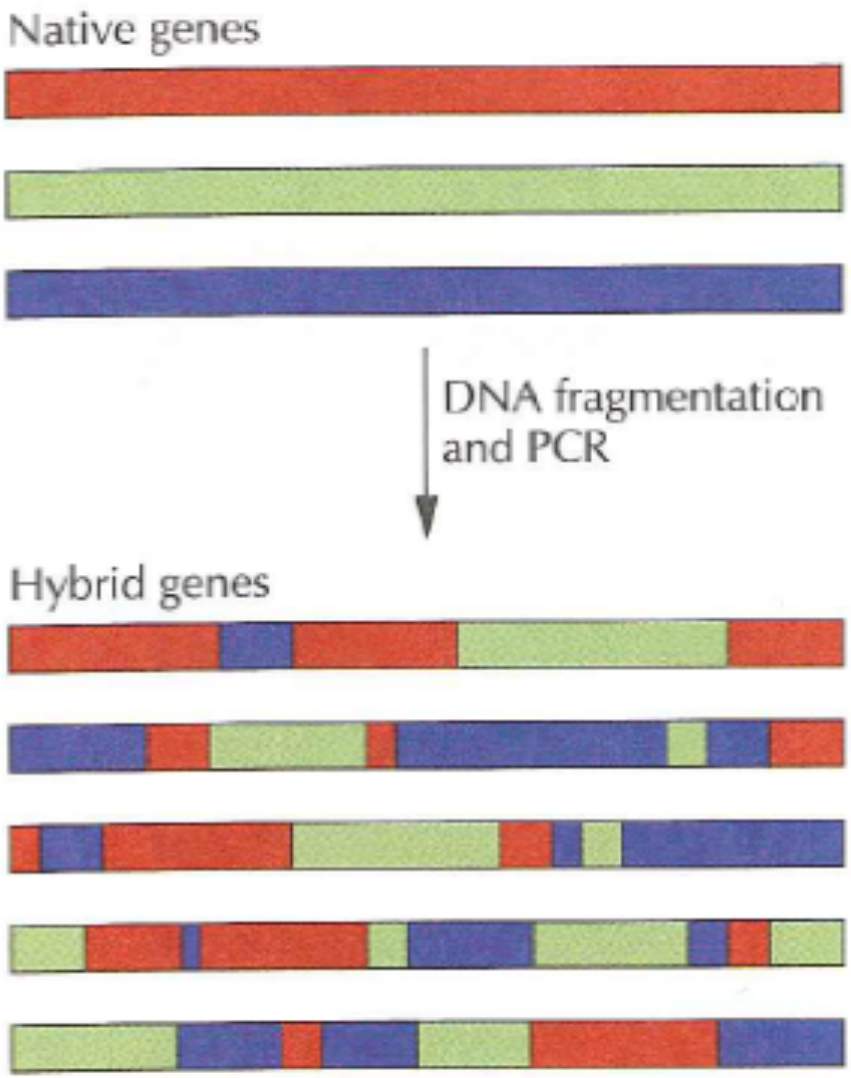
DNA shuffling - or "Sexual PCR"



Stemmer, W.P.C. *Proc. Natl. Acad. Sci USA* 91, 10747-10751 (1993).

Recombination of DNA Sequences

DNA shuffling - or "Sexual PCR"



The genes to be shuffled can be:

1. Related genes (e.g. from other organisms)
2. A parental gene that has already been mutated (e.g. by error-prone PCR)

However, homology between the genes must be ≥70% or they cannot be shuffled

It is even possible to use this technique to shuffle whole microbial genomes

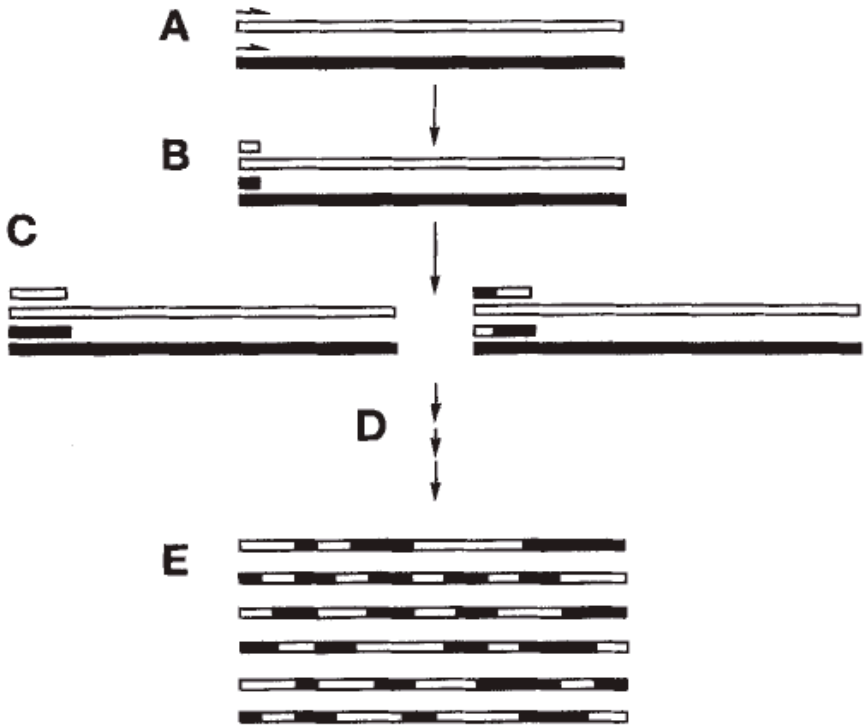
Stemmer, W.P.C. *Proc. Natl. Acad. Sci USA* 91, 10747-10751 (1993).

Recombination of DNA Sequences

Staggered extension process (StEP)

The staggered extension process (StEP) is an alternative to DNA shuffling:

A. Denatured template primed with single defined primer
B. Short fragments produce by brief polymerase-catalysed extension
C. In the next cycle fragments randomly prime the templates (template switching)
D. Repeat until full-length genes produced
E. Purify full length genes and (optionally) amplify by PCR with external primers



Zhao, H. *Nat. Biotechnol.* 16, 258-261 (1998)

Recombination of DNA Sequences

Random chimeragenesis on transient templates (RACHITT)

RACHITT allows a greater number of crossovers than DNA-shuffling or StEP

1. Create single stranded, uracil containing, bottom strand "scaffold" DNA

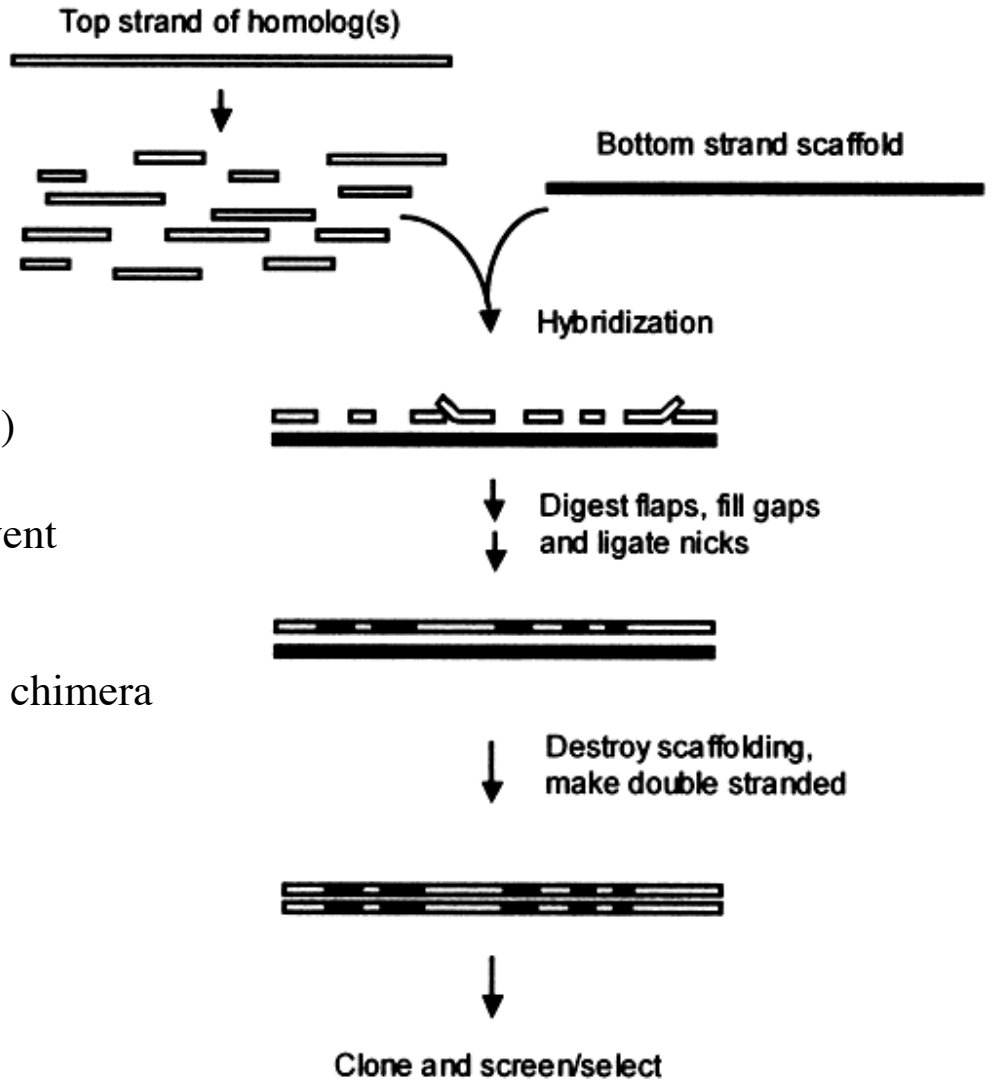
2. Hybridise DNA fragments of top strand from homologous gene(s)

3. Digest flaps (using nucleases), fill gaps (using polymerase), ligate nicks (using ligase)

4. Treat with uracil-DNA-glycosylase to prevent amplification of bottom strand

5. Amplify by PCR to create double-stranded chimera

But complicated!



Coco, W.M. *Nat. Biotechnol.* 19, 354-359 (2001)

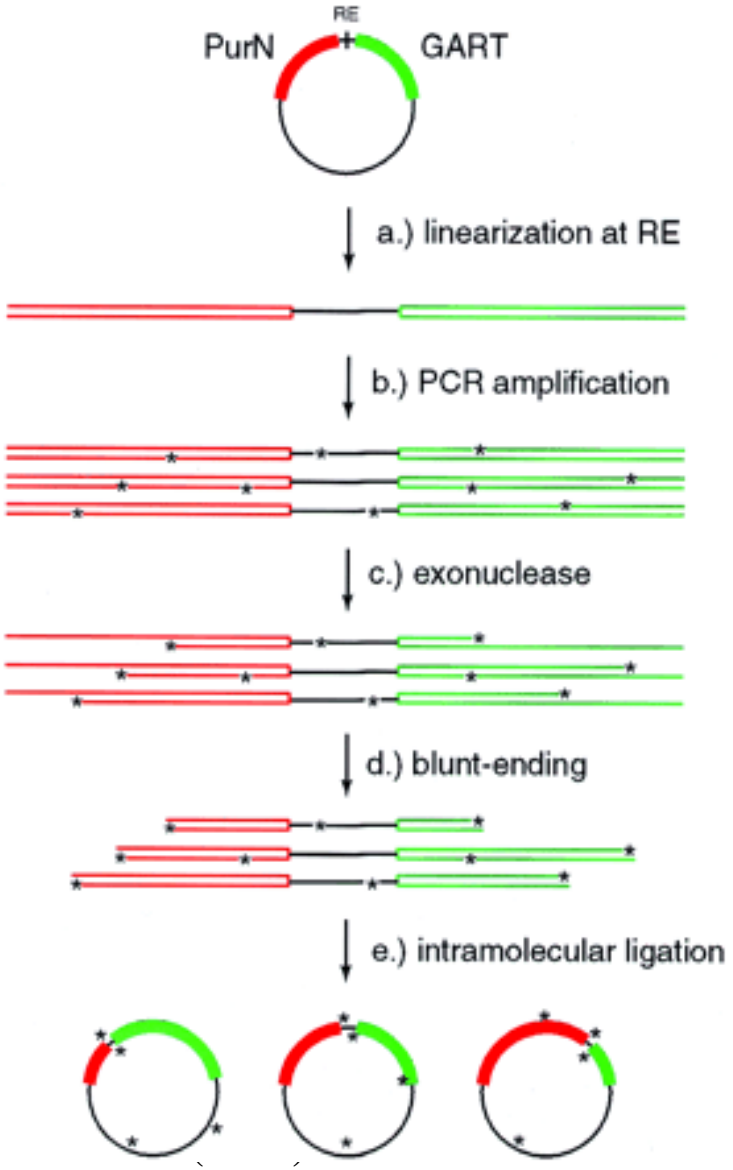
Recombination of DNA Sequences

Recombination without homology - incremental truncation for the creation of hybrid enzymes (ITCHY)

THIO-ITCHY is an improved version of the original ITCHY method.

- It allows the creation of chimeras where there is no complementarity between the genes
- However, the members of the library contain only one crossover

* Thiophosphodiester link incorporated by polymerising in presence of small amount of αS-dNTP

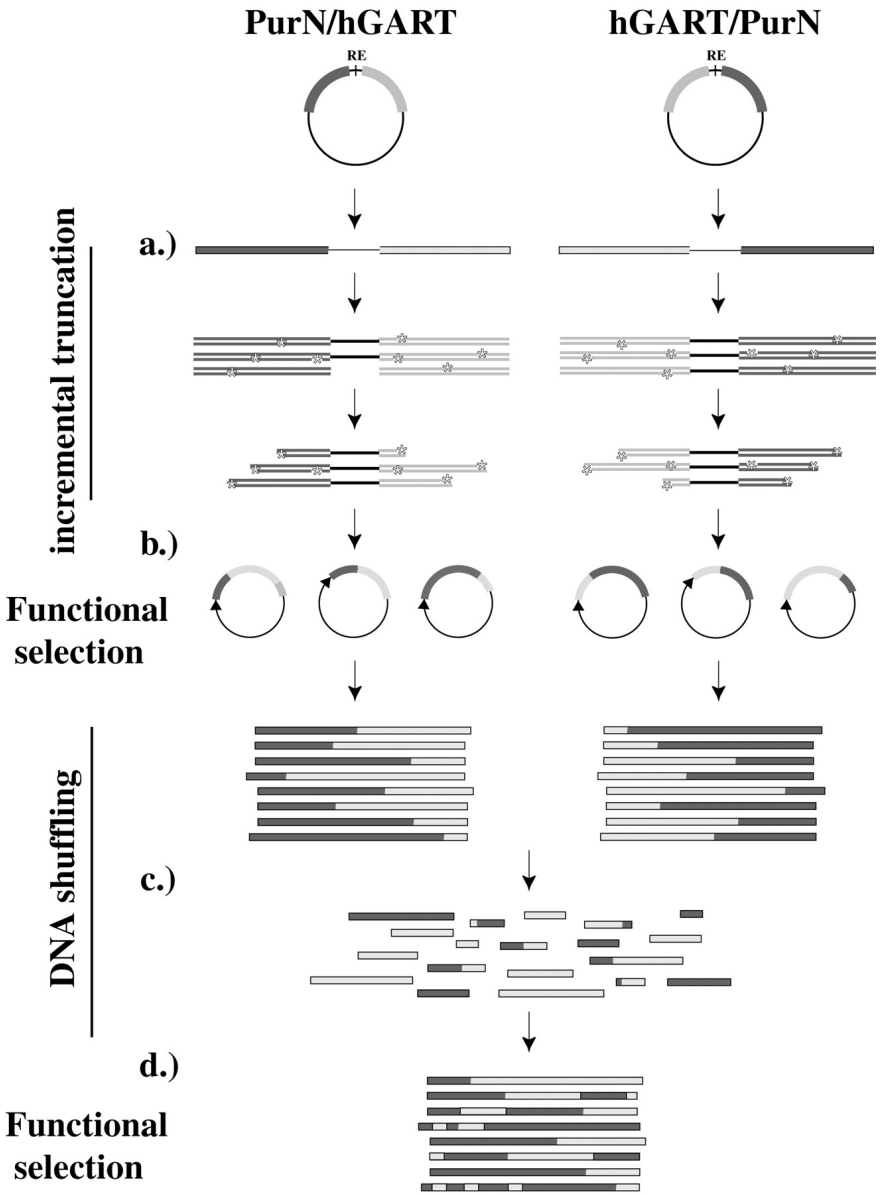


Lutz, S. et al., *Nucleic Adics Res*. 29,

Recombination of DNA Sequences

Combining ITCHY and DNA shuffling (SCRATCHY)

By combining ITCHY and DNA shuffling one can create multiple crossover DNA libraries independent of sequence identity.



Lutz, S. et al., *Proc. Natl. Acad. Sci. USA* 98, 11248-11253 (2001)