

Directed evolution of enzymes

Workshop #2 of WIDEnzymes

02.06.2025 - 06.06.2025

Department of Chemistry, Heraklion, Greece





Workshop #2: Directed evolution of enzymes



THE NOBEL PRIZE IN CHEMISTRY 2024

RATIONAL DESIGN





2. Site-directed mutagenesis



Individual mutated gene

3. Transformation

4. Protein expression

5. Protein purification

6. not applied

"for computational protein

Nobel Prize in Chemistry

2024

design and for

protein structure



Constructed mutant enzyme ©Z. Prokop &J. Damborský, Loschmidt Laboratories, Masaryk University, Brno, Czech Republic piert has received funding from the European Union's HORIZON-WIDERA-2023-ACCESS-06 programme under grapt agreement No. 101159536

IMPROVED

ENZYME



The WIDEnzymes project has received funding from the European Union's HORIZON-WIDERA-2023-ACCESS-04 programme under grant agreement No. 101159534. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

DIRECTED EVOLUTION

1. not applied





Nobel Prize in Chemistry

2018

"for the directed evolution

of enzymes"

Directed evolution

Based on the two driving forces of natural evolution:1) Differentiation through random mutagenesis

2) Selection of fitting phenotype





©Johan Jarnestad/The Royal Swedish Academy of Sciences



Directed evolution – examples from nature



Superbugs Kill More Than 30,000 Europeans Every Year

Estimated annual number of deaths related to antibioticresistant bacteria per 100,000 inhabitants, by country*



* Estimates based on data for the period 2016-2020 Source: European Centre for Disease Prevention and Control (ECDC)





THE RISE OF SUPERWEEDS

Weed species often become resistant to herbicides. Glyphosate resistance, once deemed unlikely, rose after genetically engineered crops were introduced in the mid-1990s.





Directed evolution – examples from nature



















Directed evolution







Random mutagenesis

R and a start

- No need of structure or mechanism
- Relatively easy methods of mutagenesis (UV irradiation, chemicals, error-prone PCR)

Points to consider:

- The error rate should be controlled and optimized
- Mutagenesis bias
- It is rare to get two consecutive mutations

What do we need?

- Gene of interest
- Functional expression in a (bacterial) host
- Efficient mutagenesis strategy
- Fast and accurate method of selection



And old concept – or at least theory



These results provide favourable promise for the actual construction of a machine. Its principle of operation is new - instead of trying to screen a very large number of mutants in the hope of finding advantageous offspring, one may sample the mutant distribution of say only a thousand members and probe their value topology. This value topology - explored in an iterative procedure, just as one would go about exploring the surface of a planet furnishes guidance about the best direction to continue the search, namely in the direction of highest peak density. A prerequisite of the procedure is to keep track of the branched genealogies during the successive generations of mutant spectra. The summary of our experience with RNA self-reproduction is that their optimal structures do evolve according to such a 'Darwinian logic'. (See especially Ref. 10) We can formulate the operation of this Darwinian logic as a procedure

- **10 START WITH SELECTED GENOTYPE**
- 20 LET IT REPRODUCE, MUTATING OCCASIONALLY
- **30 FORCE DIFFERENT GENOTYPES TO COMPETE**
- 40 NATURAL SELECTION OF QUASI-SPECIES AROUND BEST-ADAPTED GENOTYPE OCCURS
- **50 WHEN ADVANTAGEOUS MUTANT APPEARS GOTO 10**

M. Eigen, W. Gardiner (1984), Pure Appl. Chem. 56, 967-978.





(= evolved protein)



Directed evolution techniques





Method	Pros	Cons
Error-prone PCR	Easy to perform, mutation rate adjustable	Non-biased amino acid substitutions Only point mutations accessible
Mutator strains	Easy to perform	Entire organism/plasmid is mutated Only point mutations accessible
DNA-shuffling	Modest sequence homology sufficient Several parent genes can be used Creation of chimeras possible Useful mutations are combined, harmful ones lost	Requires sequence homology
StEP	Similar to DNA-shuffling, more simple No fragment purification necessary	Requires sequence homology PCR protocol must be specifically adapted
SHIPREC	No sequence homology required	Low diversity library in single round (might be repeated) Limited to two parents of similar length Deletions/duplications possible
ITCHY	Similar to SHIPREC	Similar to SHIPREC
THIO-ITCHY	Similar to ITCHY, but more efficient/easier	Similar to ITCHY
GSSM	All single amino acid substitutions are covered	Technically out of reach for most researchers



Methods of random mutagenesis

Error-prone PCR

- PCR with low-fidelity polymerase
- Easy method, mutations in a targeted area, easy control of mutation rate
- X Cloning issues, non-statistical change of nucleotides

DNA (gene) shuffling

- Digestion and recombination of gene fragments
- Mutations only in selected areas, cooperative effects from homologous genes
- 🗶 Reproducibility issues, cloning issues

Strain mutagenesis via chemicals or UV

- DNA damages. The DNA repair mechanisms can introduce mutations
- Easy method, about 1 mutation per 1000 nucleotides
- 🔀 Mutations are introduced all over the genome and/or plasmid





PCR









Error-prone PCR

Randomization of Genes by PCR Mutagenesis



R. Craig Cadwell and Gerald F. Joyce

Developed by Leung et al. 1989 – modified by Cadwell & Joyce 1992

Based on a PCR at conditions that increase error rate (low fidelity polymerase and non-optimal conditions)

Error-rate could be 0.5-3.0 % (or more)

False hypothesis:

All mutations happen at the same rate, thus all mutations are represented in a library equally

- Codon bias
- Polymerase bias
- Multiplication bias





Bias already was shown in the first works







FIGURE 1 Frequency of various types of mutations under the preferred reaction condition for PCR mutagenesis. (A) Mutations of the form $N \rightarrow X$; (B), Mutations of the form $X \rightarrow N$; (N = G,A,C,T; X \neq N). Frequencies refer to the mean number of mutations per base pair per PCR (30 cycles), corrected for base composition of the mutagenized gene. Error bars correspond to 95% confidence interval. Dashed horizontal line indicates expected value based on an overall error rate of 0.66% per position per PCR.

TABLE 1 Error Rate of the PCR Under Various Mutagenic Reaction Conditions^a

[dGTP] mм	[dATP] mм	Nucleotides sequenced	Mutation rate (95% C.I.)	$\frac{\text{AT}\rightarrow\text{GC}}{\text{GC}\rightarrow\text{AT}}$	Transitions Transversions	
1.0	0.2	6,001	1.37 ± 0.29%	10	2.7	
0.2	0.2	16,591	$0.66 \pm 0.13\%$	1	0.8	
0.2	1.0	1,765	$0.85 \pm 0.43\%$	2	0.4	
0.4	0.2	3,177	$0.72 \pm 0.29\%$	4	3.2	

*Reaction conditions were as described in Materials and Methods, differing only in the concentration of dGTP and dATP. Mutation rate refers to the mean number of mutations per base pair per PCR (30 cycles). Frequencies of AT \rightarrow GC and GC \rightarrow AT mutations are corrected for base composition of the mutated gene.

111 1 111 1 1 44 **** CTGGGAACTAATTTGTATGCGAAAGTATATT

1 1 11

···CAGGCATGCAC

FIGURE 3 Distribution of mutations within the gene that was amplified. (Short arrow) One occurrence; (medium-length arrow) two occurrences; (long arrow) three occurrences; no position was mutated more than three times. The DNA strand having the same sense as the RNA transcript is shown. Boxed regions correspond to primer binding sites.













Codon bias

Degenerate genetic code

- 4³ = 64 codons
- Multiple codons encode one amino acid
- Statistically, less than 6 amino acids are accessible per codon with epPCR





Lutz & Patrick (2004) Curr Opin Biotechnol 15, 291. Nevlon (2004) Nucl Acid Res 32, 1448.



Degenerate genetic code

• 4³ = 64 codons

Codon bias

- Multiple codons encode one amino acid
- Statistically, less than 6 amino acids are accessible per codon with epPCR
- Difficult to have more than one mutation per codon
- Degeneracy leads to silent mutations



Lutz & Patrick (2004) Curr Opin Biotechnol 15, 291. Nevlon (2004) Nucl Acid Res 32, 1448.





Vanhercke et al. (2005) Anal Biochem 339, 9-14

18

The WIDEnzymes project has received funding from the European Union's HORIZON-WIDERA-2023-ACCESS-04 programme under grant agreement No. 101159534. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Union European Research Executive Agency, Neither the European Union nor the granting authority can be held responsible for them.



Bias can be used to understand the copy mechanism of the lagging

Polymerases may have a preference for the bases they will mutate

- Bias can be alleviated by working with several polymerases
- and leading strands Allen et al. (2011) Nucl Acid Res 39, 7020-7033
- Under the same conditions, different polymerases may add different mutations

Polymerase bias

Funded by





DNA polymerases and fidelity







Multiplication bias



- Exponential growth on DNA products in PCR
- Over-representation of mutations that happen in the first rounds
- Up to 25% of the library can carry mutations from the first round



Possible solutions:

- Less PCR cycles, with
- More reactions with less volume, to produce more libraries



Results of bias



Methods	Transitions		Transversions				I-D-I	TIT	Deferrer	
	AT →GC	$GC \rightarrow AT$	AT →TA	AT →CG	GC →CG	GC →TA	mber	1 ₅ /1 _v	Reference	
Ideal method	16.7	16.7	16.7	16.7	16.7	16.7	0.0	0.5	Wong et al. (2006a), Wong et al. (2006b)	
Tag/MnCl ₂ and imbalanced dNTP concentration	38,1	19.0	27.8	4.0	0.0	6.3	4.8	1.5	Rasila et al. (2009)	
Taq/8-oxo-dGTP and dPTP	65.4	27.5	1.6	4.4	0.0	0.0	1.1	15.5	Rasila et al. (2009)	
Mutazyme/Amplicon	14.0	32.5	15.1	2.3	5.8	22.1	8.3	1.0	Rasila et al. (2009)	
Mutazyme/Cycle	17.1	25.7	28,6	2.9	0.0	14.3	11,4	0,9	Rasila et al. (2009)	
XL1-Red	0.0	60,0	10.0	0.0	0.0	0.0	30.0	6.0	Rasila et al. (2009)	
NH ₂ OH-HCI	15.4	76,9	7.7	0.0	0.0	0.0	0.0	12.0	Rasila et al. (2009)	
DuARChEM	13.5	26,9	1.9	5.8	42.3	9.6	0.0	0.7	Mohan and Banerjee (2008)	
Taq/D ₂ O	50.0	37.5	12.5	0.0	0.0	0.0	0.0	7.0	Minamoto et al. (2012)	
Tag/D2O and MnCl2	60.0	12,0	8.0	8.0	4.0	8.0	0.0	2,6	Minamoto et al. (2012)	
dITP and endonuclease V	38,4	39.7	4.6	8,6	4.6	4.0	0.0	3.6	Wang et al. (2013c)	
epRCA with Cre/loxP recombination	6,7	86.7	0.0	0.0	6.7	0.0	0.0	13.9	Huovinen et al. (2011)	
SeSaM-Tv -II	15.2	18.3	1.8	1.8	33.9	29.1	0.0	0.5	Mundhada et al. (2011)	
TaGTEAM	6.1	10,2	18.4	0.0	14.3	26.5	24.5	0.3	Finney-Manchester and Maheshri (2013)	



Methods to introduce mutations

- Metal concentrations affect polymerase fidelity
- **MgCl**₂: stabilizes non-complementary pairs.

It is crucial for polymerase activity

- MnCl₂: is added to reduce polymerase specificity (related to annealing temperature too)
- Some polymerases are also affected from the concentration of the template









Methods to introduce mutations



- Non-equal concentration of dNTPs
- In typical PCR they are added equally

Vartanian et al. (1996) Nucl Acids Res 24, 2627-2631.

dNTP/µN	M			Mn ²⁺	Colonies	No.				Mutation
С	Т	Α	G	mM	sequenced	mut. ^a	Ti/Tv ^b	N→A,T ^c	N→G,C ^d	frequencye
1	1000	50	50	-	37 trim ^R	126	125/1	121	5	1.5×10^{-2}
1	1000	50	50	-	24 ampi ^R	162	157/5	157	5	2.9×10^{-2}
3	1000	50	50	-	20 trim ^R	24	22/2	24	0	5.2×10^{-3}
3	1000	50	50	-	20 ampi ^R	29	27/2	28	1	6.3×10^{-3}
10	1000	50	50	-	20 trim ^R	15	12/3	12	3	3.2×10^{-3}
10	1000	50	50	-	22 ampi ^R	21	16/5	19	2	4.1×10^{-3}
5	1000	5	1000	-	18 trim ^R	3	3/0	3	0	7.2×10^{-4}
5	1000	5	1000	-	18 ampi ^R	0	0/0	0	0	$\leq 10^{-4}$
30	1000	30	1000	-	18 ampi ^R	4	4/0	1	3	10-3
30	1000	30	1000	0.5	34 ampi ^R	755	521/234	256	499	10^{-1}
100	1000	100	1000	0.5	18 trim ^R	19	16/3	8	11	4.5×10^{-3}
100	1000	100	1000	0.5	24ampiR	41	33/8	11	30	7.4×10^{-3}







Funded by the European Union

Points for consideration



- How many mutations are needed for the desired phenotype?
- What is the best template for evolution?
- Is the protein selected for evolution evolvable?
- Are negative mutations masking my beneficial mutations?
- Do I screen more one library or do I go to a second round with the best variant?
- What assay do I use for preselection?







Our workshop

Funded by the European Union



Our workshop



Masterplate







Thank you!

